Taylor & Francis
Taylor & Francis Group

Theoretical Issues in Ergonomics Science

ISSN: 1463-922X (Print) 1464-536X (Online) Journal homepage: http://www.tandfonline.com/loi/ttie20

A model for investigating cultural dimensions of communication in the car

Donal Carbaugh, Ute Winter, Elizabeth Molina-Markham, Brion van Over, Sunny Lie & Timothy Grost

To cite this article: Donal Carbaugh, Ute Winter, Elizabeth Molina-Markham, Brion van Over, Sunny Lie & Timothy Grost (2015): A model for investigating cultural dimensions of communication in the car, Theoretical Issues in Ergonomics Science, DOI: 10.1080/1463922X.2015.1107655

To link to this article: http://dx.doi.org/10.1080/1463922X.2015.1107655

THEORETICAL ISSUES IN ERGONOMICS SCIENCE, 2015
http://dx.doi.org/10.1080/1463922X.2015.1107655

Taylor & Francis
Taylor & Francis Group

A model for investigating cultural dimensions of communication in the car

Donal Carbaugh[a], Ute Winter[b], Elizabeth Molina-Markham[c], Brion van Over[d], Sunny Lie[e] and Timothy Grost[f]

[a]Department of Communication, University of Massachusetts, Amherst, MA, USA; [b]User Experience Technologies Lab, General Motors Research & Development, Herzliya, Israel; [c]Graduate Studies Office, Dartmouth College, Hanover, NH, USA; [d]Department of Communication, Manchester Community College, Manchester, CT, USA; [e]Department of Communication, St. Cloud State University, St. Cloud, MN, USA; [f]Human-Machine Interface Design, General Motors Global Vehicle Engineering, Warren, MI, USA

ABSTRACT

This article explicates a theoretical framework which was designed to discover cultural features both within and about car communication. The general approach the framework provides has been used for basic research in the United States and China, and has proven useful to researchers and designers. We discuss specifically how the approach can develop and implement speech-enabled human–machine interface (HMI) systems to address cultural features of communication and interaction. The general framework unveils the cultural nature of human–machine communication, while it also opens the possibility of discovering new cultural dimensions and principles which designers may not yet have considered. The approach is adaptable to a variety of communication contexts, with our focus here on the in-car communication of drivers with a speech-enabled HMI. Specific findings are briefly discussed including implications for research and design.

ARTICLE HISTORY
Received 2 May 2014
Accepted 9 October 2015

KEYWORDS
In-car human–machine interaction; speech system; voice-user interface; ethnography of communication; cultural discourse

Relevance to human factors

More frequently today, speech is an integral part of a driver's interface with their car; Speech allows users a natural, minimally intrusive, and efficient way of interacting with their car. However, speech itself poses a unique design challenge since, in this interface, ways of speaking are linguistically variable and culturally shaped. This sort of variation and shaping needs to be understood if users' experiences are to be satisfying and safe. This article provides a theory and methodology for empirically studying and designing this interface in ways that are sensitive to users' cultural expectations and more pleasing to them in the communication situation of the car.

1. Introduction

Extensive research has documented how the cultural nature of communication varies both within and across regions of the world (Carbaugh 1990, 2005; Hymes 1972; Philipsen

CONTACT Donal Carbaugh Carbaugh@comm.umass.edu

2002). Speech-enabled human−machine interface (HMI) design needs to take into account this variation in user understandings of and preferences for different ways of communicating in different cultural contexts. Literature provides evidence for such cultural aspects of preferred and effective user interaction, though no theoretical framework exists to formally study cultural dimensions and their design implications in the automotive environment (Tsimhoni, Winter, and Grost 2009). In this paper, our primary purpose is to discuss a framework we developed for researchers and designers to study speech-enabled HMI systems. This framework, or perspective for inquiry, is designed to discover the cultural nature of communication in contexts, while it also positions investigators to discover new cultural dimensions and principles which designers may not have considered. Although our framework can be adapted to a variety of communication contexts, we focus here on in-car communication of drivers with a speech-enabled HMI.

We begin with an overview of key concepts. Then, we introduce how these offer synergy with User-Centred HMI design principles. In the end, we demonstrate briefly the leverage our approach offers by discussing two results of our field studies concerning user interaction styles and how these resulted in re-designing HMI systems to accommodate user preferences.

We want to mention at the outset our concern about driver safety issues. A major concern of any interface design must be the maximal reduction of interference with the task of driving. Barón and Green (2006) and Peissner, Doebler, and Metze (2011) discuss the benefits of using a speech-enabled system for driving performance, but they also mention the limitations of such a system. The quality of the system's design has a substantial influence on the potential for increasing drivers' safety. Therefore, the performance of the speech recognition system is crucial (Cooper, Ingebretsen, and Strayer 2014; Kun, Paek, and Medenica 2007). Equally important is the interface design itself for keeping the complexity of today's in-vehicle applications at a safe level (Zhang and Wei 2010), thereby avoiding driver confusion during communication (Cooper, Ingebretsen, and Strayer 2014; Maciej and Vollrath 2009).

Our studies have drivers using the most user-friendly, voice-activation possible. Furthermore, all have been conducted with drivers in their own cars, driving roads familiar to them which they select. This keeps drivers in their most familiar environment both within and outside the car. The framework that follows explores how driving and voice-activation work best together (and how they do not). While we do not intend to study cognitive workload directly, we aim for the design of a culturally appropriate experience. This will contribute to users' increased feelings of intuitiveness and naturalness in the use of such systems including turn taking, interaction style, error recognition and correction, among others. Our framework thus holds potential for reducing user confusion and unnecessary complexity. The eventual contribution of our studies to safer driving will result from our field studies and data analyses presented in the following.

2. The theoretical framework

In order to conduct field research, we have developed a theoretical framework for investigating the cultural dimensions and principles of communication which influence the different degrees of success people have in dialogue with a machine. This theoretical model includes a methodology for studying in the field culturally driven user expectations, decisions and behaviours, as users interact with speech-enabled in-car systems. The general approach, the theory and the methodology, is designed to be used, and has been used in

specific regions, nations or communities, of the world (see for example Milburn 2015; Sprain and Boromisza-Habashi 2013).

2.1. The car as a communication situation

In order to understand cultural variations in the in-car HMI, we treat the car, interactions within it, and about it, as a 'communication situation'.[1] In other words, we understand the car to be a situation understood through communication, and further, we understand communication situations to be at their base, culturally variable. On these bases, we raise primary and fundamental research questions: What communication practices do people in fact do (and want to do) while in the car? Alternately, how do people talk about their car? How are these practices, and how is this talk culturally shaped and meaningful?

Our central construct, communication situation, includes several ingredients which we explore in response to these research questions: (1) the car is a place where people communicate with each other and with the car itself – in cultural ways; (2) the nature of that communication is done in ways which are distinctive to each speech community in particular; (3) those ways are structured through expressive norms, in other words, people want the interaction in, and with, the car to get done in some ways rather than in other ways; and (4) those distinctive ways, and those norms, activate users' preferences. This is a sketch of the logic in our framework, which we ground with this construct of communication situation, and which we research specifically for human–machine communication with an in-car speech-enabled HMI.[2]

2.2. In-car communication events

Within the automotive communication situation, there are specific sequences of acts that can be understood as 'communication events' (see Figure 1). In other words, each culturally situated, communication situation supports some communication events, rather than others. We understand a communication event to involve a sequence of communication

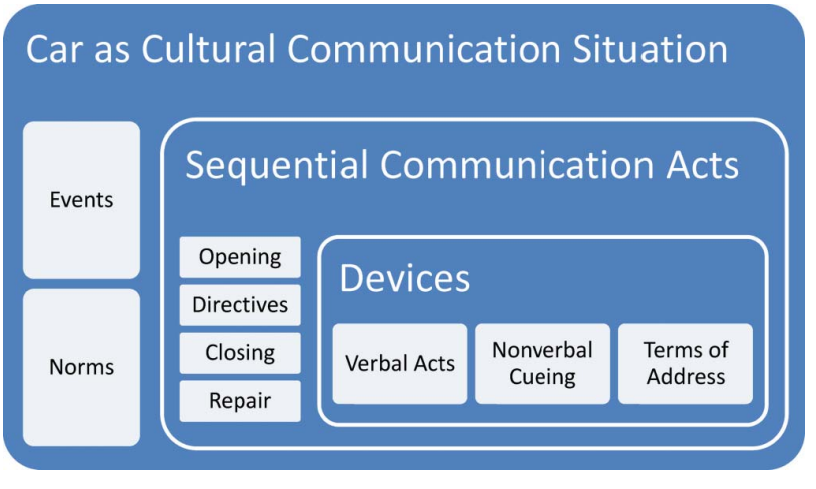

Figure 1. Central concepts in the model.

acts, which – from the participants' view – has integrity as a sequence. For example, getting the car to play music by the Beatles can be understood as a communication event, as is getting the car to identify the closest-cheapest gas station, or similarly, getting directions to that gas station, or making a phone call. Each such task requires a sequence of communication acts, which participants understand to have some degree of cultural integrity – that is, it can (and should, they think) be initiated and completed in some ways rather than others.

From the view of our framework, we are using a nested conceptualisation with communication situations including communication events which hold within them communication acts. We are developing the related point that communication acts in and about the car occur within events as 'cultural sequences'. In communication, parts of sequences are often identified by the language of participants as 'greetings', 'exchanging pleasantries', 'thanks' and the like. Each sequence has some familiar flow or some sort of integrity to participants. Knowing the cultural sequences and the flow of a particular communication event, including how it is frustrated and/or corrected, as part of the larger communication situation, can add a deeper understanding of the communication people produce in the car. This is the logic in brief of our framework.

Within a communication event there are several types of 'communication acts' we presume to be quite important to understand as humans interact in, and with a car. Some such acts are opening, directing, addressing and referencing, closing and, also, repairing trouble. All occur within the natural and routine flow of interactions and involve cultural features such as specific forms, contents and meanings.

Together with communication act sequences, our framework is designed to explore important extra-linguistic, multimodal cues. A few types of cues are gestural or tactile uses of the interface, facial expressions and other nonverbal cues. These cues can include both prosodic features (intonation, stress, pitch, register) and paralinguistic aspects (tempo, pausing and hesitation). These are purely indexical – or strictly tied to situations – because they do not necessarily have propositional content or context-free lexical meaning, but instead signal context-specific information. Different cultures have different ways of using or interpreting these 'contextualisation cues' (see Gumperz 1982). If not properly understood as such, or if not considered as part of the communication situation, such cues can lead to misunderstanding.

A key concept coupled with multimodal cues is 'conversational inference' (see Gumperz 1982, 1992; Carbaugh 2005). This concept brings into view how participants assign meaning to nonverbal cues within the on-going rhythm and flow of communication events.

2.3. Design dimensions for speech-enabled HMI

Communication practices we are discussing here give a particular understanding to practical matters of HMI interaction design – which will become clearer in our discussion below.[3] For now, consider that HMI interaction designers and researchers often develop a list of characteristics and parameters from daily work and experience by asking which of these are in question and which may vary cross-culturally. These practices or parameters are listed within our framework as dimensions of design in a way that they are observable, measurable and/or qualitatively identifiable in field data. The following summarises

several such dimensions which we have explored on the basis of our model. Each identifies potentially valuable information about cultural variability in speech-enabled HMI. Our purpose in listing these is to demonstrate how our framework embraces and develops a wide range of such dimensions:

- **User satisfaction:** What do users think of the system? What parameters, behaviours and instances lead to this opinion about the system?
- **User trust:** When and why do users trust or not trust the system?
- **Ease of use:** How well and intuitively can users understand the interface in a short time? What is easy? What is difficult?
- **User content:** How does the user find that the system deals with user content, such as contact lists, navigation data, local radio stations, etc.?
- **User interaction style:** How do users prefer to interact with the system? What is their personal communicative style? How do they expect the system to respond?
- **Multi-modal use:** Which modality do users prefer in what act or event? While driving on the road, are there things that a user feels he wants to do by touch, gestures, tactile or other modalities instead of speech? Do users feel that the visual feedback supports or is aligned with the voice interaction? What is irritating or misaligned rather than supportive?
- **Cooperative principles:** In the interactions between the user and the system, what practices make the system cooperative or not cooperative?
- **Turn taking:** Does a user know when it is his turn to interact with the system? In which situations does the user find it unexpected when the system responds or fails to respond to the user? How fast does a user take a turn? How fast should the system respond?
- **Grounding:** Do users think that the system and they have a shared understanding about the user requests and the system's capabilities to answer the requests? When do misunderstandings, conversational misalignments and points of confusion occur?
- **Conflict resolution:** How do users find the system at dealing with repairs or misunderstandings?
- **Control handling:** When do users want the system to take initiative in communication? When does the user want to be in control?
- **Information distribution:** Do users want to tell the system their request in one utterance or rather let the system guide them and ask what it needs to know, even if the dialogue is longer?
- **Audible feedback:** How do users find the wording of the responses? How do users find the quality of the system's voice?
- **Context dependency:** How does user behaviour and expectation depend on the driving situation or environment?

Interaction dimensions of design such as these, concern the verbal and/or non-verbal communication practices of the user while he/she is involved in a communication situation – or while executing a sequence of communication acts with the goal of completing a task. Each shows communicative practices that express an understanding, expectations and continuous learning of how the in-car interface works and communicates back. Therefore, while observing and analysing the user's communication practices in the car, we can learn about the user's expectations, seeing in an exacting way dimensions of interface design. In the end, as a result of these studied observations, we

can derive design recommendations for interfaces in the in-car communication situation. Thus, we ground all design questions and areas in the framework of the larger communication situation, with special attention to specific communication events and acts. With this focus, we are attentive to local norms or preferences which we theorise as follows.

2.4. Norms in communication situations and events

As we have already noted, communication situations and events have been defined as the activities, or aspects of activities, that are directly governed by rules or norms for the use of speech. The concept of communication event draws attention to ways participants go about doing each interaction sequence and emphasises that each can be done in a proper way according to participants' preferred conduct within the car.

Norms are 'statements about conduct which are granted some degree of legitimacy by participants' (Carbaugh 2007, 178). They are messages about 'correctness' that may be stated explicitly by participants or they may be more implicit as a dimension in the structure of participants' practices. In our view, a norm includes the essential element of an 'ought' or a 'should'. In this sense, communication norms tap into a moral domain of practical action, practices which people believe should be conducted in some ways, rather than in others. This way of understanding communication norms, then, is not simply a description of a regular or normal routine (i.e., the thing people routinely do), but of a normative practice (i.e., what people believe should or should not be done).

Communication norms, as practical actions, can of course vary in their force (see Philipsen, Coutu, and Covarrubias 2005; cf. Jackson 1965, 1975). As we adapt Jackson's early ideas, the concepts of intensity and of crystallisation help distinguish the variety of dimensions in and types of communication norms.

- **Intensity:** How strongly do people feel about the normative practice? On a 1−7 Likert scale , a bi-polar distribution with half at 6−7 strongly liking it, and half at 1−2 strongly disliking it may be possible. The intensity, then, regarding this norm is strong, with respondents registering the strength of their feeling at the extreme ends of the scale. Of course, if most respondents were in the 3−5 range, we discover that people report not feeling very strongly about this norm, or it is rather weak in its intensity.
- **Crystallisation:** Through such analyses, we can know further if there is or is not general agreement with regard to this practice of communication. When there is a bi-polar distribution in the results − as when some favour it while others do not – the norm would have, as discussed earlier, a high degree of intensity since people feel relatively strongly about it, yet a low degree of crystallisation since people do not agree on how it should be done.

Again as we adapt Jackson's earlier ideas, there are three types of communication norms which derive from these dimensions of intensity and crystallisation:

(1) **Conflict Potential:** These are norms with high intensity and low crystallisation.
(2) **Normative Power:** These are norms with high intensity and high crystallisation.
(3) **Vacuous Consensus:** These are norms with low intensity and high crystallisation.

Knowing what the topical matters of concern are for participants and which norms are associated with these matters, with what degree of importance, enables us to design

a communication environment which people feel is more their own, closer to the expectations and desires they hold. We refine our understanding of these situations by examining particular communication events including the norms or rules active within them.

The key objectives then, of the framework, are several: (1) to discover specifically how in-car communication gets done; (2) to describe HMIs generally, as communication situations which are culturally distinct and therefore cross-culturally diverse; (3) to interpret participants' meanings of communication in these situations; (4) to study comparatively the nature of communication in these situations; and (5) to provide guidance in the design of communication in these situations.

We now discuss our methodology, generally. Our purpose here is to cast our treatment of the methodology as unattached to a particular field site thereby demonstrating its utility in any field site around the globe. We do note the framework and methodology have been fruitfully used and developed in the United States and in China (see Milburn 2015; Molina-Markham et al. in press; Sprain and Boromisza-Habashi 2013). Our treatment of the methodology is in two sections, the first focused on overall field design and data collection, the second on three phases of qualitative data analysis.

3. Research design: data collection in the field

Since human–machine interaction depends to a large extent on the communication situation, its communication acts, events and norms, we consequently need to create an environment in which users are able to act as natural as possible if we want to observe their practices, in this case, in the car environment. Any field study design must, therefore, be guided by the principle of 'ethnographic or naturalistic inquiry', observing users in their natural car environment, driving where each wants, acting as each would like and interacting with the car and its HMI in their preferred ways. Our general methodology involves a sequential research design in four general phases, with each phase involving a specific set of activities from the researchers' activities prior to entering the field, to activities completed after leaving the field (see Figure 2).

Pre-fieldwork Activity: This phase of a project involves several activities, which are preliminary to doing the fieldwork itself. It is crucial to be as knowledgeable about a field site as possible. If there is a literature available about the site, then it can and

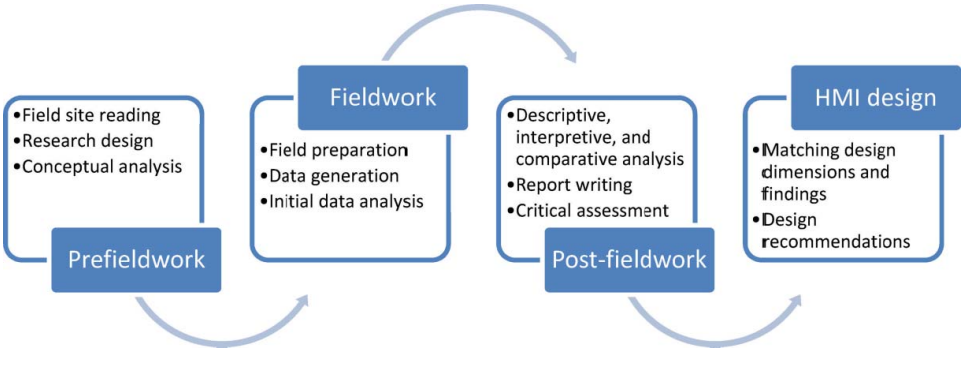

Figure 2. A brief portrayal of the field methodology.

should be consulted. What is the history of the area, its people, its economy, driving tendencies, occupations and so on? It is helpful to have some local knowledge about where one is going and who are the participants in the field study. A second set of activities involves preliminary planning about the fieldwork itself. According to the special focus envisioned, a preliminary conceptual map can and should be formulated. Focusing on specific concerns establishes a theoretical position from which to observe communication in (and about) the car. This equips one for study and reflection while in the field, enables a systematic approach to one's observations, as well as provides ways of designing interviews (and eventually cars). We emphasise that this conceptualisation – the framework and interview protocol – is not a closed design, to go unrevised after one is in the field, but an open, yet structured design to be reflected upon and revised as needed based upon subsequent field activities.

Fieldwork Activity: This phase of activity involves periods of observation, in this case of communication in the car itself. Observations such as these inevitably lead to subsequent questions that researchers want to ask participants. The accumulation of a descriptive record about observational and interview events creates a large corpus of data, which then needs to be analysed. The analysis involves the distinct modes of investigation discussed in the following section.

Post-fieldwork Activity (A): During this phase, researchers conduct detailed descriptive analyses first, and then, conduct detailed interpretive analyses. The activities conducted after leaving the field involve deeper phases of analysis. These analyses lead, often, to additional questions about the dialogic in-car dynamics observed, or heard about while in the field. This phase of the research can lead, when possible, back to the field for more detailed observations and analyses. It can also lead to further studies, which are conducted under modified conditions, for example in a controlled experiment focussing on predefined parameters for observation, or in a simulated driving environment.

Post-fieldwork Activity (B): Conduct comparative analyses, and critical assessments; write empirical field research reports; offer interface design and policy recommendations based upon findings. These final phases of a field project are crucial as each creates a sharper view of the cultural dimensions of communication getting done in, and about the car. These are often better understood through comparative analysis, for instance, by conducting the field studies in different cultures. Note, and we emphasise, that the phases we present here are sequential in their design, but cyclical in their possibilities. A phase of fieldwork can result in revising one's earlier conceptualisations; post-fieldwork activities can lead one back to the field focused on other observations, with different questions; writing a report can lead to its presentation, with that presentational event being an occasion for subsequent observing and interviewing; and so on. The general research design is linear but its typical implementation is cyclical (see Carbaugh and Hastings 1992).

3.1. Field data analyses: descriptive, interpretive, comparative

In our discussion earlier, we have discussed one key purpose of our framework, collecting data in the field. Here, we address three – descriptive, interpretive and comparative - ways those data can be subsequently analysed.

The above framework focuses on the car as a communication situation, and within it, specific events and acts which are culturally shaped. But how are these to be analysed? Here we provide a set of components for such descriptive analyses. For such descriptive analyses, there are eight basic components we use to analyse a communication situation, event or act (see Hymes 1972; Carbaugh 2012). We will use the concept, communication practice, here, as an umbrella construct, which includes the concepts above, such as communication situation, communication event, communication act, and all of the dimensions we discuss of the HMI design. Following the central objectives of the framework we discussed earlier, the following components are used initially to describe the communication that is happening, and then subsequently to interpret the meanings of that communication to participants. The components also provide a basic general set of elements for comparing communication practices across communities, nations or scenes.

For an initial **descriptive analysis**, the components function as a series of questions to be asked about communication practice, thereby discovering how communication is done within and about the car. We summarise this descriptive use of the components as follows:

(1) **Setting**: In what physical environment is the communication taking place?
(2) **Participants**: Who is involved in the communication practice?
(3) **Ends**: This component has two parts: What are the participant's goals of the practice (e.g., to send an email)? What are the outcomes of the practice (e.g., the email was sent, or the effort to do so was unsuccessful, or the user got irritated at the car)?
(4) **Act/Sequences**: What specific communication acts are getting done, and in what sequence?
(5) **Key**: What is the emotional pitch, or tone of the communication (e.g., perfunctory, serious, frustrated)?
(6) **Instrumentalities**: What multiple mode(s) or cues is being used in this communication (e.g., voice, gesture, pressing or touching a button)?
(7) **Norms** (see last section): What are the norms – stated and/or implied – for this interaction?
(8) **Genre**: Is there a generic form to this communication practice which participants use, and if so, what is it?

These eight components provide a basic investigative framework for analysts to systematically describe any communication practice, such as communication in and about the car.[4]

Our general methodology is designed in distinct phases. A second phase of analysis focuses on **interpretive inquiry**. During this phase, the above framework (along with an additional set of analytic procedures) will again be used, but this time to interpret the participants' meanings of the communication practice described earlier. In other words, the same concepts can be used as follows:

(1) **Scene**: What do participants say is 'going on' in this setting, through this communication practice?
(2) **Participant identities**: What culturally significant roles, identities or relationships are active in this practice, from the participants' view?
(3) **Ends**: What are the goals and outcomes sought by participants in this practice, in their own terms?
(4) **Act/Sequences**: Is there a regular routine, and perhaps an ideal routine, or act/sequence for doing this? What is it? How 'relaxed' or strictly formulaic is that sequence?
(5) **Key**: How do participants interpret the feeling of this practice?

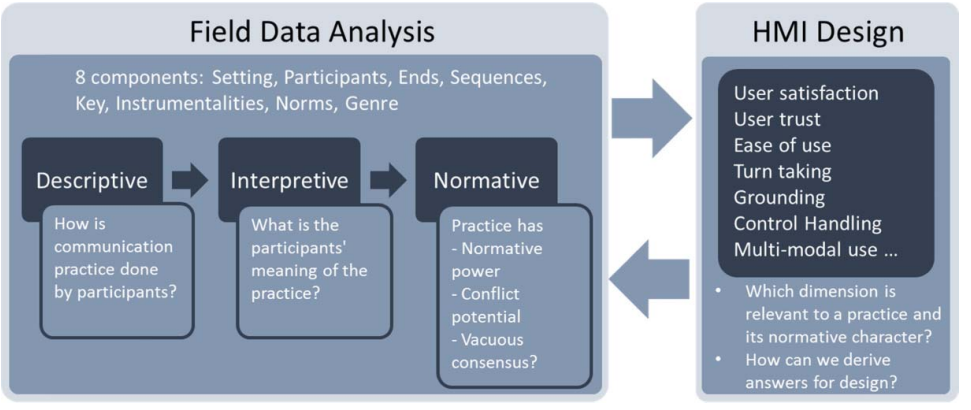

Figure 3. Analysis method of field data.

(6) **Instruments**: What are the necessary, and/or preferred modes or channels for doing this activity?
(7) **Norms**: What norms do participants employ for interpreting this activity?
(8) **Genre**: Is there a generic cultural form active here, and can it be identified by participants?

In this second step, as in the descriptive analysis, the HMI design questions will be consulted to extract HMI-relevant insights from the data such as user interaction styles, and the preferences associated with those insights, which we discuss later in the article (see Figure 3).

The eight components in relation to the HMI design dimensions are being used here, then, in two ways; first, to describe the activities in the car mainly from an analyst's point-of-view; then to investigate, and unveil the participants' meanings of those communication practices, in that situation, through their sense of the events, or through their ways of speaking. Consequently, the results will lead to design recommendations for the in-car speech-enabled HMI.

A following, third phase, of comparative analyses respond to the questions: to what degree is this practice the same, here and there, and to what extent is it different? If one were to study direction-giving in Massachusetts and in Shanghai, or speech prompting of entertainment, we would expect some similarities and some differences in this practice in the two locations. Through such study, we get a better idea about what indeed is culturally distinctive in one set of practices, as opposed to those in another; we also get a better sense of what is similar across such practices. Through such comparative study, and based upon descriptive and interpretive analyses, we are able to determine what parts in the speech-enabled HMI need a different design, and what parts can display the same behaviour in all examined cultures. While it is obvious that due to different languages the system prompts need to be transferred from one culture to the other, there may be less evident considerations related to dialogue flow, task directives, addressing, interaction style, communication tone, user trust and others, which may have influence on the prompt design for each culture. For example, in our initial comparative studies, we have discovered less complex dialogue flow in directives (i.e., fewer communication acts or shorter sequences) in our Chinese data than in our US data.

Figure 4. User perspective of embedded Wizard of Oz HMI.

3.2. Experimental set-up for field research about in-car communication

3.2.1. An embedded Wizard of Oz interface for the car

For our empirical studies, we replaced a user's in-car infotainment system with an embedded Wizard of Oz multimodal interface, providing the user with four typical domains for infotainment: phone dialling, radio tuning, music selection and navigation. Passonneau et al. (2011) gives an overview on the use of embedded wizardry for dialogue systems including a brief discussion of system accuracy when a human wizard replaces part of the recognition process. Figure 4 shows our system mounted on the air vent in the car centre console and a view of a driver.

The application allows the user two modes of interaction. The user can choose to conduct all tasks via touch screen in the centre console, when field study regulations permit. This kind of interaction will be handled by the application without intervention from the Wizard. The user may also choose to initiate a task event by speech after touching a speech icon on the screen. In this case, the task of the Wizard is to replace the Speech Recognizer and Natural Language Understanding module, transforming the user intention into a touch sequence on his Wizard screen, including the choice of confidence levels that a machine would produce.

Using such an embedded Wizard system encourages the user to behave most naturally and in the user's preferred way, while the front-end and dialogue management of the system is a machine and perceived as such by the user (see Figure 5).

3.2.2. Multimodal design principles

The embedded Wizard system allows us to explore design principles, which cannot yet be implemented to this degree in speech applications due to technological limitations. Jurafsky and Martin (2009) or Jokinen and McTear (2009) give an overview of the status of speech technologies and dialogue systems.

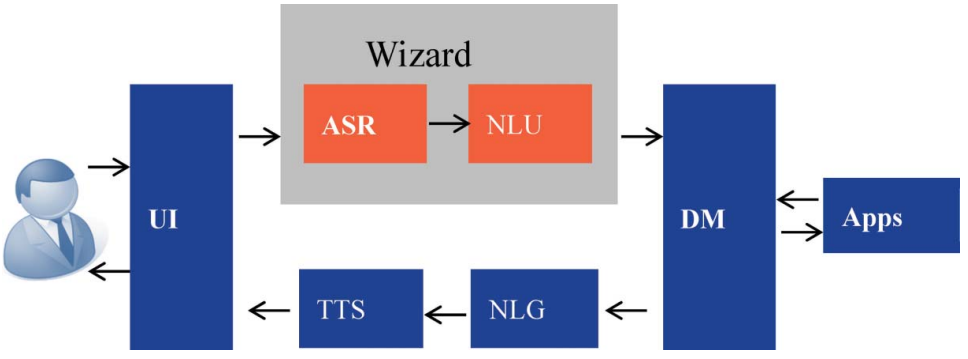

Figure 5. Illustration of embedded wizardry into dialogue system.

As for spoken dialogue sequences, the Wizard system supports full Natural Language Understanding, flexible information distribution within a dialogue sequence, mixed-initiative dialogues, context dependency and multimodal support. Consequently, the Wizard of Oz interface is highly cooperative. The user can take a turn at almost any point during a dialogue, he can interrupt the system, change his request, express himself freely, etc. Moreover, the Wizard can learn about user preferences and context during the study and use such knowledge to adapt the system's understanding to the user.

As for manual–visual interactions with the touch screen, the user interface relies on several design principles. Visual cues, such as colour, icons, location and spacing/grouping elements, are used with the aim of improving the level of feedback to the user. According to the principle of cross-modal priming, visual prompts echo spoken prompts, but the visual display is also 'stand alone'. Another design principle is the use of shaping text. Once the user presses the speech button, example sentences appear to shape the spoken interaction and provide information of how to use the system. Finally, vertical and horizontal sets of tabs serve to provide visual context for speech communication.

The Wizard of Oz system is a fully symmetric multimodal infotainment system that integrates both visual and manual capabilities, as well as speech input and output throughout a flexible interactive experience. The system allows switching between modalities at any point of an interaction sequence. We discuss next one way the above theory and methodology was applied.

4. One brief example of a field data collection process

One of our field projects 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 typically experience a high volume of traffic. Twenty-six participants were selected so as to represent a wide variety of characteristics, 12 male and 14 female drivers. Nineteen of the participants were originally from the Northeastern United States. Participants ranged in age from 26 to 64 years 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 9 hours driving per week. Nineteen of the participants were smartphone users, and seven were not. Participants were recruited through flyers

posted in public areas or through snowball sampling procedures as a result of direct contact with researchers. The overall session with each participant lasted about 120−150 minutes. (Our other field sites were in urban Shanghai and Beijing, China, with the same research approach being used in both.)

Participants brought their own car to the study, as well as their personal content to the Wizard of Oz interface, such as the address book of their mobile phone or their preferred music collection. As two team members installed the Wizard system and observation equipment for visual and audible documentation (which records the participant's face, the tablet screen, the dashboard of the car and the view out of the front windshield), the participant was introduced to the experiment and given the consent form. The field research team typically included two researchers in the car: the 'experimenter' in the front passenger seat and one researcher, the 'Wizard', operating the Wizard of Oz interface in the back seat. A third researcher helped with the equipment installation.

Once in the car, the participant was introduced to the basic capabilities of the infotainment system and was then invited to explore interacting with the system through visual−manual means and specifically through speech, while experimenting with using the system for a variety of tasks. Then the participant drove around a large, open, off-road area until comfortable with the tested interface. Afterwards the participant performed a 60−75 minute drive of their choosing. During the drive, the participant was encouraged to use the Wizard of Oz interface in his preferred way with the exception of not using touch, other than touching the speech button or ending an application.

After completing the drive, the driver was asked a series of questions based upon our interview guide. The explored dimensions of the dialogue experience, including turn taking, task completion, error correction, quality of audible feedback, multi-modal use, general ease of use and user satisfaction, among others. All questions were open-ended questions (Winter, Tsimhoni, and Grost 2011).

Finally, the researchers met together in a debriefing session to reflect upon the specifics of the driving session, the interview session, particular observations made during each, and to identify useful focal concerns for further attention and analysis.

4.1. From field data findings to interface design: one general recommendation for user interaction style

The following section gives a brief demonstration of how we have used our analyses based upon collected field data for in-car HMI design. In the initial analysis of our findings from the field studies, we analysed participant practices related to communication events such as task initiation and switching between tasks, directives or commands to the in-car system, variety in the cultural formulation of directives, task-ending events and misalignments and misunderstandings (Molina-Markham et al. 2015). In this section, we briefly preview two very general findings and the implications these hold for one aspect of HMI design. The presentation should not be understood as a full qualitative and quantitative detailed analysis because it serves solely for demonstrating the validity and practicability of our approach.

4.2. Analysis of observable user interaction styles

Our field data included these communication events:

Participant 20
21:34.74

1. ((radio playing; participant touches microphone button and system dings))
2. System: What sports channel do you want to hear?
3. Participant: Go to NPR.
4. System: Hold on. ((pause)) Tuning radio to 88.5 FM WFCR.
5. ((radio station changes))

Participant 12
01:11:45.40

1. ((XM radio playing; participant touches microphone button and system dings))
2. System: What XM channel?
3. Participant: Call H-.
4. System: Got it. ((radio stops playing and phone rings)) calling H- at home.

Our data also included multiple instances like the following:

Participant 13
01:18.85

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

Participant 9
53:32.70

1. Participant: ((touches microphone button and system dings)) Give me WFM-uhhhh
2. give me W:: what's it called FCR
3. System: Pardon?
4. Participant: FCR
5. System: Please confirm 88.5 FM WFCR
6. Participant: Good. ((pause)) Good job ((shows thumb up))

7. System: Tuning radio to 88.5 FM WFCR ((radio station plays))

The first two instances show one sort of preference from a user; the second two indicate another. As we analysed our data, we wondered how to characterise these sorts of communication events, the preferences they exhibit for HMI design, including asking participants what they prefer in such interactions with the in-car system.

In this regard, in our descriptive and interpretive analyses, we focused primarily on questions related to the components of Act/Sequences, Ends, Participant identities, Key and Norms. In terms of Act/Sequences, we were interested in how the sequences unfolded and how some instances (e.g., the last two presented here) included more acts than others. Considering the component of Ends, we noted that participant goals as far as accomplishing tasks and relating with the system seemed to vary among instances. In the last two examples there seems to be a goal of being entertained through the dialogue or establishing companionship with the voice in addition to the task completion of the HMI system. This notion of relating with the system introduced questions about Participant identities in terms of how participants understood their relationship with the system in their car. In contrast to the sense of companionship in the last two instances, the first two instances show participants relating to the voice as an assistant or secretary with a distant or neutral relationship.

These observations are supported by participants' interview data in which they expressed their needs and expectations of the HMI system. Participants with preferences similar to the first two examples preferred a quick, efficient exchange. Even when asked about their practices, these participants gave rather short answers. For example, participant 12 answered 'Yeah that's totally ok', when asked if she likes the prompts. In contrast, participant 13 expressed that 'it would be nicer if it kind of understood my conversational style well enough' and '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'. Participant 9 even said: 'Well if I say "I wanna hear some wacko political rhetoric" and you don't know what that means… and of course as soon as I hear that stuff I wanna talk about it with somebody so… it's tempting not to wanna talk back to it'. Participant 24 expressed the importance for a more conversational style as follows: 'you feel like you're interacting with somebody and having a conversation', and 'that's something that I just feel better doing that, I have no idea why. It's very strange. When I think about it, it just feels more natural'.

Certain norms of interaction guide the progression of the act sequence, the types of goals that participants believe can and should be accomplished through these sequences, and the relationships between users and a machine that participants believe should or should not exist in this context. Regarding Norms, we have seen that the field data reveals different norms pertaining to interaction styles. We found two different normative preferences for styles.

4.2.1. A style of maximal efficiency

One way of using the system is to get the job done, the task completed, as quickly and efficiently as possible. In this style, the above features configure, in an extreme form, in this way: task initiation is done in a one-shot manner (e.g., 'play radio 88.5'), switching between tasks is done in the same way (e.g., 'now call John Smith at home'), elaborations to directive formulations are absent or minimised, tasks are ended by one-shot voice

commands or a touch of the end button. Some participants expressed an interest in using this style as the predominant style.

4.2.2. A style of maximal interactivity

A second way of using the system is different. Participants like playing with it, calling it a name, inviting it to respond to them (and desiring that). In this style, the above features and dimensions configure, in an extreme form, in this way: task initiation is prefaced with extra-task talk (e.g., 'well I wonder what I am in the mood for right now…'), as is switching between tasks (e.g., 'well, should I call Bill now or Janice…'), directives or commands to the in-car system are made at times by naming the system (e.g., 'OK Eliott, what's it gonna be now…") or with an endearment term (e.g., 'nice job Darlin"), variety and elaboration is evident in the cultural formulation of directives including several communication acts (e.g., 'how about some country'; 'do we have any Garth Brooks available...'; 'is his latest CD in there…'; 'what about the duet with Patty Loveless?' and so on), and task ending events include some politeness features such as 'thanks' and 'see you again'. Some participants expressed a desire for such a style as a predominant one.

The previous analysis included 22 participants (participants 4−25), because participants 0−3 participated in a pilot data collection period and experienced various difficulties; they did not use the final experimental set-up. Among the 22 participants, there were three predominantly interactive users (participants 7, 13, 24), and three mix-style users who would use either efficient or interactive styles (participants 9, 18, 19). All the other participants were predominantly efficient users, some of them with an occasional tendency for an interactive utterance (participants 12, 14, 15, 17, 20, 21, 25). Our analyses suggest, at this time, that many if not most participants would enjoy having both styles available, perhaps among others, so as to choose what suits their mood at a particular time or in a particular driving session.

Also as a preliminary observation, we find that our Chinese data include more preference for efficiency than interactivity. This demonstration is further developed elsewhere (Carbaugh 2012; Molina-Markham et al. 2015; Wang, Winter, and Grost 2015; Winter, Shmueli, and Grost 2013).

4.3. Initial considerations for in-car HMI design recommendations

As people use their in-car communication system, at times they prefer a style of maximal efficiency. When they do, they want to minimise the time it takes for dialogue and the number of turns it takes to complete a task. In our data, this sort of preference has a high degree of intensity. The implications for interaction design are as follows: create a possibility for the interaction flow to be as concise and efficient as possible, subordinate prompt design to this optimisation principle, remove unnecessary prompts or other elements in the dialogue process, in grounding, and in turn taking. When the maximal efficiency protocol is active, users find the absence of interactive elements to be acceptable and even pleasant. And in turn, when the optimal efficiency style is preferred, the interactive elements in the prompts are treated at best indifferently or at worst as quite irritating.

Users at other times prefer a style of maximal interactivity. When this is the case, users show a preference for variation, relational talk, naturalness and being entertained. At these

times, users appear to want to develop a personal relationship with the system. This preference seems to be guided by a norm about the importance of being pleasant. The implications for interaction design in this case are that prompts should be designed to maximise their perceived pleasantness rather than the efficiency of the interaction flow. Thus, prompts for this situation should be designed to progress toward task completion, but the priority is on elements that are perceived as pleasant.

In the light of these findings, we conclude that a system that can determine user interaction style and adapt to this style would be ideal. A set of system prompts could be designed that accommodate each of the two styles that we have identified here. With this type of a system, users could select which style they prefer and could also potentially switch between styles. One way of accommodating this switching would be to first design prompts emphasising the efficiency principle and then develop these prompts by adding interactive elements, so that the basic prompt is the same for both sets of prompts.

5. Summary

How a person chooses to interact with the various forms of technology that surround him or her is deeply connected to the cultural context in which those interactions take place. In this paper, we have explicated a theoretical framework and its associated methodology for conducting empirical field studies that explore the cultural dimensions of individuals' communication with speech-enabled interfaces in their car. This model includes key concepts from the ethnography of communication and cultural discourse theory including communication situation, event and rules or norms. These are used in conjunction with several design dimensions to produce HMI-relevant findings regarding HMI design dimensions. In this tradition the experimental set-up and analyses are designed to reveal qualitative as well as quantitative results. We briefly demonstrated one productive finding of our research concerning one HMI dimension, styles of user interaction.

One finding from our field studies includes these two different styles – one with a focus on efficiency in task completion and the other that emphasises elaborate interactivity and a personal relationship with the system – that are preferred by participants to varying degrees in a northeast region of the United States. We add at this point that our field data collection in China, and its comparative analyses, revealed the same two general interaction styles, with differences between the two mainly being in the language itself including the dialogue flow relative to its sequential structuring. This reveals differences in directive formulation, error corrections and event structure generally. Using these styles as guides, among other culturally distinct features (see this research teams' other published field studies), researchers can develop speech-enabled systems that more closely align with user expectations and preferences, here demonstrated briefly to illustrate interaction flows and prompt design. In terms of future work, the field data we have collected is invaluable, deeply rich and detailed enough to facilitate the development of the exact phrasing of prompts, the detailed specification of the voice persona for both types of interaction styles, the flow and rhythm of the system, as well as development of the other design dimensions discussed earlier.

Acknowledgements

The development of the theory and methodology reported here was supported by a research grant from General Motors Research and Development to the first author. The first author is appreciative of this support, to the delightful collaborative work with Dr. Ute Winter, and to our considerable research team, some of the members listed here as co-authors and others with General Motors in Israel, Michigan and Shanghai.

Notes

1. The model derives from a long tradition in studies of cultural dimensions of communication from the seminal conceptualization of Hymes (1972), through Philipsen (1987, 2002), to Carbaugh (1988, 2007), among others.
2. Our recent studies explicate different aspects of the methodology (Carbaugh et al. 2012, 2013).
3. The design of an in-vehicle multimodal interface traditionally follows the paradigm of User-Centered Design (Norman and Draper 1986; Nielsen 1993; Vredenburg and Butler 1996), which is widely considered the key to product usefulness and usability (Mao et al. 2005). Naturalistic observation, contextual inquiry and other techniques (e.g., Holtzblatt 2003; Holtzblatt, Wendell, and Wood 2004; Dray and Siegel 2007) are used to learn about user preferences in the relevant context, for example, the driving situation and environment.
4. A sub-set of the concepts may be more useful in some cases than in others. In other words, not every component has equal importance or value in every investigation. These eight concepts are inclusive of all of the design dimensions that we have identified earlier. In conducting the descriptive analysis, we analyse each practice through the eight concepts in general and subsequently for all of the detailed HMI dimensions. Because of the applied method of naturalistic inquiry, we explore each of the detailed HMI design dimensions through these concepts, keeping open the possibility of discovering the unknown, for example, the design dimension which has normative power in a culture, but which we have not thought about or observed in other cultures yet.

Disclosure statement

No potential conflict of interest was reported by the authors.

About the authors

Donal Carbaugh is a professor of communication at the University of Massachusetts Amherst and author of the resent book, *Reporting Cultures on 60 Minutes* (Taylor and Francis).

Ute Winter is with the User Experience Technologies Lab of General Motors Research and Development, Herzilya, Israel.

Elizabeth Molina-Markham is an independent researcher, Boston, Massachusetts, was post-doctoral Researcher on this project and specializes in cultural approaches to communication.

Brion van Over is an assistant professor of communication at Manchester Community College, Connecticut who specializes in cultural features of conversational interaction.

Sunny Lie is an assistant professor of communication at California Polytechnic University Pomona with specialties in intercultural interactions.

Timothy Grost works in human machine interface design at General Motors Global Vehicle Engineering in Warren, Michigan.

References

Barón, A., and P. Green. 2006. *Safety and Usability of Speech Interfaces for In-vehicle Tasks While Driving: A Brief Literature Review*. Technical Report UMTRI-2006-5, University of Michigan Transportation Research Institute Library.

Carbaugh, D. 1988. *Talking American: Cultural Discourses on DONAHUE*. Norwood, NJ: Ablex.

Carbaugh, D. 1990. *Cultural Communication and Intercultural Contact*. New York, NY: Erlbaum Publishers.

Carbaugh, D. 2005. *Cultures in Conversation*. Mahwah, NJ: Routledge.

Carbaugh, D. 2007. "Cultural Discourse Analysis: Communication Practices and Intercultural Encounters." *Journal of Intercultural Communication Research* 36 (3), 167–182.

Carbaugh, D. 2012. "A Communication Theory of Culture." In *Inter/Cultural Communication: Representation and Construction of Culture*, edited by A. Kurylo, 69–87. Thousand Oaks, CA: Sage.

Carbaugh, D., and S.O. Hastings. 1992. "A Role for Communication Theory in Ethnography and Cultural Analysis." *Communication Theory* 2 (2): 156–165.

Carbaugh, D., E. Molina-Markham, B. van Over, and U. Winter. 2012. "Using Communication Research for Cultural Variability in Human Factor Design." In *Advances in Human Aspects of Road and Rail Transportation*, edited by N. Stanton, 176–185. Boca Raton, FL: CRC Press.

Carbaugh, D., U. Winter, B. van Over, E. Molina-Markham, and S. Lie. 2013. "Cultural Analyses of In-car Communication." *Journal of Applied Communication Research* 41 (2): 195–201.

Cooper, J.M., H. Ingebretsen, and D.L. Strayer. 2014. *Mental Workload of Common Voice-based Vehicle Interactions Across Six Different Vehicle Systems*. Washington, DC: AAA Foundation for Traffic Safety.

Dray, S., and D. Siegel. 2007. *Understanding Users in Context: An In-depth Introduction to Fieldwork for User Centered Design*. Vol. 4663 of *Human-Computer Interaction – INTERACT. Lecture Notes in Computer Science*, edited by *C. Baranauskas et al*., 712–713. Berlin, Heidelberg: Springer.

Gumperz, J.J. 1982. *Discourse strategies*. Cambridge: Cambridge University Press.

Gumperz, J.J. 1992. "Contextualization and Understanding." In *Rethinking Context: Language as an Interactive Phenomenon*, edited by A. Duranti and C. Goodwin, 229–252. Cambridge: Cambridge University Press.

Holtzblatt, K. 2003. "Contextual Design." In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, edited by J. Jacko and A. Sears, 941–963. Mahwah, NJ: Lawrence Erlbaum Associates.

Holztblatt, K., J.B. Wendell, and S. Wood. 2004. *Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design*. San Francisco, CA: Morgan Kaufmann.

Hymes, D. 1972. "Models for the Interaction of Language and Social Life." In *Directions in Sociolinguistics: The Ethnography of Communication*, edited by J.J. Gumperz and D. Hymes, 35–71. New York: Basil Blackwell.

Jackson, J.M. 1965. "Structural Characteristics of Norms." In *Current Studies in Social Psychology*, edited by I.D. Steiner and M. Fishbein, 301–309. New York: Holt, Rinehart & Winston.

Jackson, J.M. 1975. "Normative Power and Conflict Potential." *Sociological Methods and Research* 4: 237–263.

Jokinen, K., and M. McTear. 2009. "Spoken Dialogue Systems." Vol. 5, *Spoken Dialogue Systems: Synthesis Lectures on Human Language Technologies*, edited by G. Hirst. Morgan and Claypool. doi: 10.2200/S00204ED1V01Y200910HLT005

Jurafsky, D., and J.H. Martin. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. 2nd ed. Upper Saddle River, NJ: Prentice-Hall.

Kun, A., T. Paek, and Z. Medenica. 2007. "The Effect of Speech Interface Accuracy on Driving Performance." In *Proceedings of the 8th Annual Conference of the International Speech Communication Association (Interspeech '07)*, Antwerp, Belgium.

Maciej, J., and M. Vollrath. 2009. "Comparison of Manual vs. Speech Based Interaction with In-vehicle Information Systems." *Accident Analysis & Prevention* 41 (5): 924–930.

Mao, J., K. Vredenburg, P. Smith, and T. Carey. 2005. "The State of User-Centered Design Practice." *Communications of the ACM* 48 (3): 105–109.

Milburn, T., ed. 2015. *Communicating User Experience*. London: Lexington Books.

Molina-Markham, E., B. van Over, S. Lie, and D. Carbaugh (2015). " "OK, talk to you later": Practices of Ending and Switching Tasks in Interactions with an in-car Voice Enabled Interface." In *Communicating User Experience: Applying Local Strategies Research to Digital Media Design*, edited by T. Milburn, 7–25. London: Lexington Books.

Molina-Markham, E., B. van Over, S. Lie, and D. Carbaugh. In press. "You can do it baby": Cultural Norms of Directive Sequences with an In-Car Speech System. *Communication Quarterly*.

Nielsen, J. 1993. *Usability Engineering*. San Francisco, CA: Morgan Kaufmann.

Norman, D., and S. Draper. 1986. *User Centered System Design; New Perspectives on Human-Computer Interaction*. Mahwah, NJ: Lawrence Erlbaum Associates.

Passonneau, R.J., S.L. Epstein, T. Ligorio, and J. Gordon. 2011. "Embedded Wizardry." In *Proceedings of the SIGDIAL 2011 Conference*, 248–258. Portland, Oregon, June 17–18, 2011.

Peissner, M., V. Doebler, and F. Metze. 2011. "Can Voice Interaction Help Reduce the Level of Distraction and Prevent Accidents?" In *Meta-Study Driver Distraction Voice Interaction*. Pittsburgh, PA. Carnegie Mellon University, White Paper.

Philipsen, G. 1987. "The Prospect for Cultural Communication." In *Communication Theory: Eastern and Western Perspectives*, edited by D.L. Kincaid, 245–254. San Diego, CA: Academic Press.

Philipsen, G. 2002. "Cultural Communication." In *Handbook of International and Intercultural Communication*, edited by W. Gudykunst and B. Mody, 51–67. London: Sage.

Philipsen, G., L.M. Coutu, and P. Covarrubias. 2005. "Speech Codes Theory: Restatement, Revisions, and Response to Criticisms." In *Theorizing about Intercultural Communication*, edited by W.B. Gudykunst, 55–68. Thousand Oaks, CA: Sage.

Sprain, L., and D. Boromisza-Habashi. 2013. "The Ethnographer of Communication at the Table: Building Cultural Competence, Designing Strategic Action." *Journal of Applied Communication Research* 41: 181–187.

Tsimhoni, O., U. Winter, and T. Grost. 2009. "Cultural Considerations for the Design of Automotive Speech Applications." In *Proceedings of the 17*th *World Congress on Ergonomics IEA 2009*, Beijing, China.

Vredenburg, K., and M. Butler. 1996. "Current Practice and Future Directions in User-Centered Design. In *Proceedings of Usability Professionals' Association Fifth Annual Conference*, Copper Mountain, CO.

Wang, P., U. Winter, and T. Grost. 2015. "Cross Cultural Comparison of Users' Barge-in with the In-Vehicle Speech System." *Design, User Experience, and Usability: Interactive Experience Design*. Vol. 9188 of *the Series Lecture Notes in Computer Science*, 529–540. Cham, Switzerland: Springer International.

Winter, U., O. Tsimhoni, and T. Grost. 2011. "Identifying Cultural Aspects in Use of In-vehicle Speech Applications." In *Proceedings of the Afeka AVIOS 2011 Speech Processing Conference*, Tel Aviv, Israel.

Winter, U., Y. Shmueli, and T. Grost. 2013. "Interaction Styles in Use of Automotive Interfaces." In *Proceedings of the Afeka AVIOS 2013 Speech Processing Conference*, Tel Aviv, Israel.

Zhang, H., and L.N. Wei. 2010. "Speech Recognition Interface Design for in-vehicle System." In *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (AutomotiveUI '10). New York: ACM.