



Adding Communicative and Affective Strategies to an Embodied Conversational Agent to Enhance Second Language Learners' Willingness to Communicate

Emmanuel Ayedoun¹  · Yuki Hayashi¹ · Kazuhisa Seta¹

Published online: 30 July 2018
© The Author(s) 2018

Abstract This paper describes an embodied conversational agent enhanced with specific conversational strategies aiming to foster learners' readiness towards communication in a second language (L2). Willingness to communicate (WTC) in a second language is believed to have a direct and sustained influence on learners' actual usage frequency of the target language. To help overcome the lack of suitable environments for increasing L2 learners' WTC, our approach is to build embodied conversational agents that can help learners surmount their apprehension towards communication in L2. Here, we focus on the dialogue management aspects of our approach and propose a model based on a set of communication strategies (CS) and affective backchannels (AB) to foster such agents' ability to carry on natural and WTC-friendly conversations with learners. We examined learners' expected WTC after interacting with one of the following versions of the system: an agent featuring both CS and AB; an agent featuring only CS; and an agent featuring only AB. The results suggested that combining CS and AB empowers the conversational agent and leads to higher expected WTC among L2 learners. We also found that even the AB-only version of the system had the potential to enhance WTC to some extent. These findings are evidence of the feasibility of enhancing L2 learners' engagement towards communication using a computer-based environment coupled with appropriate conversational strategies.

Keywords Willingness to communicate in L2 · Embodied conversational agents · Communication strategies · Affective backchannels

✉ Emmanuel Ayedoun
eayedoun@ksm.kis.osakafu-u.ac.jp

¹ Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University, Sakai, Japan

Introduction

The ability to communicate in a second language (L2) is a significant asset in facilitating exchanges between people who come from different countries or speak different languages. One of the fundamental goals is to provide learners with the ability to communicate effectively in the second language whenever the opportunity presents itself. The key factor to ensuring such communicative readiness is the willingness to communicate (WTC), defined as the “readiness to enter into discourse at a particular time with a specific person or persons, using an L2” (MacIntyre et al. 1998). Following the finding that learners with a higher WTC tend to perform better than others in producing the target language, MacIntyre et al. (1998) suggested that increasing learners’ WTC should be the goal of L2 learning. Moreover, they proposed a pyramidal heuristic model of variables affecting WTC, in which it appears that the environment where learners experience or practice the second language plays an important role in motivating them to take part (or not) in conversation. However, many learners feel genuine anxiety about performing in front of others, and many classrooms do not offer learners much in the way of communicative practice (Reinders and Wattana 2014).

The goal of this study is to contribute to enhancing L2 learners’ willingness to communicate by providing them opportunities to freely simulate and enjoy immersive daily conversations in a computer-based conversational environment. Here, L2 communication is problematic, because it involves the learners’ ability to communicate within the restrictions of their vocabulary, grammar, etc. Thus, unlike communication between native language (L1) speakers, breakdowns or pitfalls in communication occur more often with L2. Therefore, any conversational agent intended to support communication in L2 should adopt strategies adapted to such interactions.

In this paper, we propose a Dialogue Management model based on Communication strategies and Affective backchannels (DiMaCA), a model based on a set of specific conversational strategies (i.e., communication strategies (CS) and affective backchannels (AB)) dedicated to fostering conversational agents’ ability to carry on WTC-effective conversations with learners in an English-as-a-foreign-language (EFL) context. Here, we report on the practical significance of using the above-mentioned conversational strategies to enhance L2 learners’ WTC.

The paper is organized as follows. We begin with a brief review of past and present studies related to in L2, conversational strategies in L2 communication, and spoken dialogue systems. Next, we outline the novelty and main contribution of our work. Then, we provide an overview of the set of conversational strategies (i.e., CS and AB) employed in this study and the resulting dialogue management model (i.e., DiMaCA) and its characteristics. Later, we describe the pilot study, its procedures, results, and implications. Finally, we present some concluding remarks and discuss directions for future work.

Literature Review

Willingness to Communicate in Second Language Learning

The use of the target language plays a crucial role in L2 acquisition (Seliger 1977). Some researchers have emphasized the role of output (i.e., the use of the target

language) in L2 learning, arguing that it is necessary for the development of production (i.e., speaking and writing skills). This is because input (i.e., the process of understanding what is said or written) develops only listening and reading comprehension (Swain and Lapkin 1995; Swain 1998). However, a pressing issue is how to encourage learners to use L2 for communication, because many students do not naturally engage in much L2 production, either inside or outside the classroom. Moreover, some L2 learners, despite excellent linguistic competence, tend to avoid using L2 for communication, whereas others with only minimal linguistic competence seem to communicate via L2 whenever possible. Such differences can be explained by the fact that the intention or willingness to engage in L2 communication, rather than linguistic competence, is determined by a combination of immediate precursors, such as learners' perception of their own L2 proficiency (i.e., perceived competence or self-confidence), desire to use the language in a specific context, and lack of apprehension about speaking (i.e., L2 anxiety) (MacIntyre et al. 1998). Many other studies have confirmed a positive correlation between a combination of a lower level of anxiety with a sufficient level of perceived competence and willingness to communicate in a second language, suggesting that learners who experience a lower level of communication anxiety and display a higher perceived communicative competence tend to be more willing to use the second language in communicative situations (Clément et al. 2003; Compton 2004).

Following these findings, many researchers from different countries, such as Yashima in Japan (Yashima 2002), Oz in Turkey (Öz et al. 2015), and Peng in China (Peng 2007), have intensively investigated the validity of the WTC model in their own respective contexts. Although some differences, owing to each country's cultural and social characteristics, may exist, it is generally acknowledged that the variables identified by MacIntyre et al. constitute a basic and universal reference model of key factors influencing WTC in L2. Additionally, studies have shown that learners displaying high WTC are more likely to practice (MacIntyre et al. 2001), show more improvement in their communication skills (Yashima et al. 2004), use L2 in authentic communication (Kang 2005), and acquire higher levels of language fluency (Derwing et al. 2008).

To foster L2 learners' WTC and encourage meaningful interaction during class, researchers have cited important things teachers should do, such as lowering students' anxiety, making the lesson topic interesting and relevant, facilitating student acceptance of the necessity to use the target language for communication, and instilling positive attitudes towards the international community, including interest in international affairs, willingness to go overseas to stay or work, readiness to interact with intercultural partners, etc. (Yashima 2002). Notably, whereas research into computer-mediated communication in the context of second language acquisition has proliferated in the past two decades, only a few studies have investigated practical ways to enhance L2 learners' WTC. Compton (2004), for example, revealed that chatting helped students feel confident and, consequently, willing to participate orally in classroom discussions. However, he also reported that its impact on WTC varied from learner to learner and was dependent on several factors, including the topic of discussion and the attitudes of partners. Nakaya and Murota (2013) developed a mobile conversation learning system, aimed at motivating Japanese learners to communicate in English. In their system, conversation topics were based on learners' lifelogs (mainly daily activities or events posted by learners on their social network account) or related to situations that learners

often experience in daily life. Still, the “conversations” were mainly system-driven, so that learners were limited to answering questions generated without any possibility for them to get help from the system when they faced difficulty answering questions. Therefore, conversational breakdowns consistently occurred during L2 conversations with this system.

In our previous work (Ayedoun et al. 2016), we proposed an embodied conversational agent (ECA) based on MacIntyre et al. (1998)’s model to help increase second language learners’ willingness to communicate by providing opportunities to naturally simulate daily conversations in various social contexts. Our evaluation demonstrated the system’s potential to simulate natural conversations in a specific context and the feasibility of using a computer-based environment for improving learners’ WTC, even though the observed “WTC gains” among learners were not statistically important. Our results were also consistent with prior observations by Reinders and Wattana (2014). Their study employed an online multiplayer game to provide Thai learners with opportunities to chat and talk with other players in English while playing the game. The results indicated that providing such contextual conversation opportunities to L2 learners led to an increase in their engagement towards communication in L2. Of course, conversation opportunities alone are not enough to sustainably motivate learners towards communication in L2; it requires explicit care with the variables affecting learners’ WTC. Therefore, we hypothesized that a good level of conversation smoothness, to be achieved by implementing strategies to keep the conversation going, may help learners to face difficulties by creating a friendly conversational environment where learners feel at ease (Ayedoun et al. 2016). As pointed out by Mesgarshahr and Abdollahzadeh (2014), language learners, especially those at lower levels, are likely to experience difficulty when communicating in the target language. They added that too much difficulty during communication might cause them to abort their attempt and may result in them having a diminished desire to communicate. Thus, the ability to help learners overcome difficulties when communicating in L2 should be considered essential for any system intended to increase WTC among L2 learners.

Communication Strategies

When human-to-human communication is disrupted, the speaker will likely make an effort to resolve the conversational trouble by employing a variety of strategies and tactics (Long 1981). Even nonverbal actions can be a way of solving a communication problem. There are several such communication strategies (CSs). For example, *approximation* is a CS that consists of using a term that expresses the meaning of the target lexical item as closely as possible (e.g., saying “the thing you open bottles with” instead of “corkscrew”). CSs were derived from the notion of “problem-orientedness”. In other words, they were viewed as devices to compensate for communication gaps between a speaker and a listener. For a number of years, they were objects of intensive research in second language acquisition (Faerch and Kasper 1983). The taxonomies for CSs vary depending on the researcher’s approach and the type of research being conducted. The early research dealt with different languages, different levels of learners, and different procedures. It has been suggested that the differences in target item, proficiency level, and learner type resulted in different CS selections (Bialystok 1983; Ito 2000; Liskin-Gasparro 1996; Poulisse 1987). Thus, different variables affect

the selection of CSs. Therefore, it is necessary to consider variables that potentially affect the results of research and the taxonomy that best fits one's research design.

Dörnyei and Thurrell (1991) referred to the ability to appropriately use CSs as strategic competence. They considered to be a component of communicative competence (Canale and Swain 1980). It is conceivable that underdevelopment of this competence may account for some L2 learners' lack of ability to overcome interactional pitfalls, which may adversely affect their WTC. Nevertheless, in the case of learners having a low WTC, achievement of communicative competence does not automatically guarantee L2 usage (Mesgarshahr and Abdollahzadeh 2014). This is because WTC in L2 is directly affected by other variables such as anxiety and self-confidence. These variables affect psychological preparedness to communicate at a given moment. It might, therefore, be necessary for the dialogue partner (e.g., the conversational agent) to provide frequent feedback to reassure the learner and encourage him/her to pursue the interaction.

Backchannels

Listeners' behaviors in relation to their interlocutors may affect L2 speakers' fluency during oral tasks. Such behaviors have received much deliberation (Wolf 2008). When L2 speakers perform oral tasks, teachers and/or testers are often present, and they respond to the production with a variety of verbal and nonverbal messages. In the literature, such messages have been variously described as "signals of attention" (Fries 1952), "accompaniment signals" (Kendon 1967), "listener responses" (Dittmann and Llewellyn 1968), and "backchannels" (Yngve 1970), among several others. It is generally thought that, people in conversation often convey information through two channels: a predominant or main channel, through which speech flows, and a secondary, or backchannel, dedicated to sending meta-conversational signals (White 1989). Backchannels should be more properly understood as verbal or non-verbal expressions occurring in a conversation's secondary channel; they serve a meta-conversational purpose in the sense that they do not bring any new content-based information to the communication. Rather, they are used to support the primary speaker's turn by conveying the listener's comprehension and/or interest.

Moreover, although most verbal backchannels are brief utterances, even short questions, statements, and sentence completions may also be regarded as backchannels (Yngve 1970; Duncan and Fiske 2015). Additionally, researchers have found that backchannels may have interactive functions within discourse, including a regulatory function (Schegloff 1982), a repair function, and a clarification function (Hayashi and Hayashi 1991).

These "extra-linguistic" signals play a powerful role in defining the nature of social exchange. When the signals are positive, they can lead to feelings of rapport and promote beneficial outcomes in diverse circumstances, e.g., negotiation and conflict resolution (Drolet and Morris 1998; Goldberg 2005). Previous studies have amply demonstrated the importance of such backchannels during human-agent conversations, considering them an important milestone for building engaging and natural virtual humans (Kopp et al. 2008; Morency et al. 2010). These studies provide insight on the idea that backchannels may support affective variables influencing L2 learners' WTC in a conversational agent-mediated interaction.

Spoken Dialogue Systems

Spoken dialogue systems are defined as computer systems with which humans interact on a turn-by-turn basis, and with which spoken natural language plays an important role in communication (Fraser 1997). Such systems can be broadly divided into two categories: domain-specific task-oriented systems designed to assist users to achieve goals in specific domains (Bohus and Rudnicky 2009) and open-domain non-task-oriented systems, which aim to entertain through open-ended chatting with users (Banchs and Li 2012). Considering that task-oriented systems have a clear potential for computer-assisted language-learning systems that place the student in a realistic situation, where a task should be accomplished in the target language (Raux and Eskenazi 2004), we decided to focus on building a conversational agent can engage in domain-specific task-oriented dialogue with L2 learners.

Recent technological progress has led to the appearance of humanoid interfaces, or embodied conversational agents (ECAs). An ECA is a computer-generated animated character that can carry on natural, human-like communication with users (Cassell et al. 2000). ECAs add a social dimension to the human-machine interaction, increasing the believability of agents and intensifying the user's feeling of engagement with the system (Van Mulken et al. 1998). Social agency theory (Atkinson et al. 2005) outlines the effectiveness of animated pedagogical agents in human-computer interaction. According to this theory, using verbal and visual cues in a computer-based environment encourages learners to interpret their interaction with the computer as a partnership. Learners consider their interaction with the computer to be a social one, because the social cues are similar to what they would expect from a human-to-human conversation.

A rich variety of studies have employed ECAs as tutors to help users learn a particular task (Rickel and Johnson 1998; André et al. 1998), virtual guides to provide information about expositions to visitors (Kopp et al. 2005), pedagogical agents to teach users technical or scientific topics (Lester et al. 1997), etc. According to Bälter et al. (2005), "*virtual tutors can teach in ways that are impossible in the real world*". For instance, they can use animations to externalize or display features that a human teacher would not be able to show. For instance, ARTUR, a speech training system with articulation correction (Bälter et al. 2005) uses three-dimensional animations of the face and internal parts of the mouth (tongue, palate, jaw, etc.) to display phonetic features that would be hidden inside a human speaker in order to give feedback on the difference between the user's pronunciation and the correct one. Moreover, ECAs' effectiveness in educational settings has been studied as it relates to teaching science, mathematics, and humanities (Arslan-Ari 2010). For instance, animated pedagogical agents have been shown to offer L2 learners the opportunity to interact with "native speakers" and to provide a social context (Ohmaye 1998). A good example illustrating the benefits of such agents is the Tactical Iraqi Language and Culture Training System. It is an advanced computer agent-based software program, which was initially developed to give US Marines practical training in Iraqi culture, gestures, and situational language skills prior to their deployment on real-world missions. The system employs advanced artificial intelligence techniques to place Marines in a virtual-world computer-game environment where they must perform face-to-face communication tasks with animated characters representing local people (Johnson and Valente 2008).

Much of the recent research on second-language acquisition has taken a cognitive approach, as suggested by VanPatten and Benati (2015). The cognitive approach deals with the processes in the brain that underpin language acquisition, for example, how paying attention to language affects the ability to learn it, or how language acquisition is related to short-term and long-term memory. Similarly, the above-mentioned Tactical Iraqi Language and Culture Training System, as well as most other tutoring systems, tend to embody cognitive learning theories, such as skill-based theories of second language acquisition, to help learners acquire language proficiency and maximize learning outcomes. These theories imply that “language is viewed as a skill that is learned through practice, which provides opportunities for developing declarative knowledge (e.g.: knowledge of rules for grammatical accuracy) into procedural knowledge (e.g.: knowledge about how to use the grammar to speak accurately) as language use become more automatic” (Chapelle 2009, p.747).

However, our interest here is rooted in the emotional processes affecting learners’ motivation towards L2 usage. Thus, beyond the cognitive dimension of L2 learning processes, we aim to build ECAs that are fully dedicated to raising learners’ engagement towards L2 communication per their affective and behavioral state.

In the field of second language acquisition, more specifically when considering support for such emotional variables affecting language acquisition and usage, there has been limited research on the use of ECAs in multimedia learning environments. As it stands, whereas some conversational agents help simulate authentic interactions, they are not aimed at L2 learners’ willingness to communicate. We could not find any that were specifically designed for, or suitable for, improving it.

Contribution and Novelty of this Study

In light of the findings and limitations of the different studies mentioned above, it becomes clear that, most of the significant proposals to make learners more willing to communicate in L2 have come from the fields of communication or language learning. In the areas of computer-assisted language learning or artificial intelligence in education, the topic seems to be a conspicuous rarity in the literature, because traditional spoken dialogue frameworks seem to not consider aspects of L2 learners’ WTC. Although there have been a few attempts (Compton (2004), Nakaya and Murota (2013), Reinders and Wattana (2014)) at devising computer-based approaches to increase L2 WTC, not much effort has been expended on investigating the usage of realistic virtual interfaces, such as ECAs, which seem to have the potential to be a suitable alternative to face-to-face spoken interactions. Building on our previous work (Ayedoun et al. 2016), in which we showed that a dialogue agent-based conversational environment might be effective in increasing L2 learners’ WTC, we here propose DiMaCA, a dialogue management model dedicated to facilitating the implementation of ECAs that helps to raise learners’ engagement. The originality of our approach lies in its usage of two conversational strategies (i.e., CS and AB), which allows it to consider both aspects related to communicative breakdowns that often occur in L2 learner-agent interactions and those related to affective variables influencing L2 WTC, in accordance with MacIntyre’s WTC model. By enabling the ECA to make use of CS, our idea is to enhance its own strategic competence to release learners from the challenging and WTC-inhibiting burden of resolving communication pitfalls by

themselves. By identifying a novel category of backchannels (i.e., AB), we aim to foster the ECA's ability to convey empathetic and WTC-friendly support to learners.

Researchers (Nass et al. 1994; Bickmore and Picard 2005; Park and Catrambone 2007) have amply demonstrated that the human-computer relationship is fundamentally social, suggesting that the insights and lessons learned from face-to-face interactions and the above-mentioned language learning theories might also be applicable to agent-human learning situations. However, it is still unclear whether all learners react to computer agents as they would to human partners. Thus, through this research, we aim not only to enhance L2 learners' engagement towards communication with a computer agent-based system, but to also collect quantitative and qualitative data about the relationship between conversational agents and L2 learners, which might be useful for developing a generic model of such interactions.

Conversational Strategies to Increase WTC

Proposed Dialogue Management Model

DiMaCA aims to use CSs to foster ECAs' ability to handle their own difficulties and learners' pitfalls by making possible relatively smooth interactions between L2 learners and dialogue agents, whereby learners' confidence increases. Second, by way of ABs, this model aims to achieve warm interactions where learners feel less anxious about L2 communication. Hence, the rationale of implementing DiMaCA can be explained in terms of increasing L2 learners' confidence via CSs and reducing their level of anxiety towards communication via ABs. As well, previous studies on predictors of L2 learners' WTC (MacIntyre and Charos 1996; MacIntyre et al. 1998) showed a positive correlation between the decrease in anxiety and the rise of confidence towards communication among L2 learners. Therefore, we cannot deny that ABs, by reducing L2 learners' anxiety, might also have an effect of increasing confidence and vice-versa.

In the following paragraphs, we explain the characteristics of the strategies used in DiMaCA.

Communication Strategies

CSs occur on the predominant or main channel of communication; they are “a systematic technique employed by a speaker to express his or her meaning when faced with some difficulty” (Dörnyei and Scott 1997). These difficulties can arise either from a speaker lacking linguistic resources, or from an interlocutor who cannot understand the speaker. It is worthwhile for learners to have a repertoire of such strategies at their disposal, whereby they can achieve a degree of communicative effectiveness beyond their immediate linguistic means (Thornbury 2005). Nevertheless, in the case of learners with a low WTC, mastering such strategies does not necessarily guarantee that they will be able to use them when they face trouble during a conversation. In contrast, the use of CSs by ECAs can help such systems overcome their own difficulties, mainly in regard to understanding the learner on one hand, and giving answers to his/her requests on the other, as we will describe in the following section. More importantly, the use of CSs by ECAs can also help them handle communication pitfalls

(e.g., learners' difficulty to understand the agent's utterances or answer to the agent's requests) that learners may encounter during interactions. We hypothesize that learners who feel that they can rely on a supportive dialogue agent to help them recover from difficulties may develop a "sense of security" that will increase their confidence towards communication in the target language. In the present study, we were confronted to the lack of prior research work/evidence on the usage of communication strategies for improving L2 learners' WTC, as well as examples of their explicit implementation in ECAs. However, instead of defining such strategies from scratch, we targeted nine suitable ones among those defined in the comprehensive review of definitions and taxonomies of communication strategies (Dörnyei and Scott 1997). The selected strategies were chosen according to two criteria: (i) their probable usefulness for encouraging WTC from a heuristic standpoint (i.e., relying on empirical rules of judgment, such as the probable WTC-friendliness of the selected strategy); and (ii) the feasibility of their implementation from the technical standpoint. Based on these criteria, for example, a CS such as *Response reject* (i.e., rejecting what the interlocutor has said or suggested without offering an alternative solution), even though it is technically implementable, was not selected because it cannot be considered a WTC-friendly strategy. Table 1 lists the selected strategies and examples of their use in our system.

Affective Backchannels

Backchannels are verbal or non-verbal expressions given by the listener on a conversation's secondary channel (i.e., they do not bring any new content to the conversation) to show interest, attention, or willingness to keep the channels open. They play a major role in human-agent conversation (Smith et al. 2011). Although actual competence might encourage communication, it is the perception of competence that ultimately determines the L2 learner's choice to communicate (Clément et al. 2003). Moreover, the degree of attention from others that L2 learners get might have an influence on the apprehension they feel towards communicating (McCroskey 1997). Thus, it is conceivable that L2 learners who do not get enough supportive feedback from their interlocutors perceive themselves as being incompetent communicators and therefore, tend to be reticent to communicate. It follows that it might be effective for a conversational agent intending to enhance learners' WTC to convey enough interest or sympathy to learners at times during the interaction. Doing so might reduce anxiety and encourage learners to take risks in using the target language. To achieve such empathetic support, we identified a set of affective backchannels to enable ECAs to explicitly convey thoughtful support to learners by congratulating them, cheering them up, or showing sympathy in accordance with the interaction state. Table 2 shows the four different categories of ABs that we introduce to help the conversational agent provide L2 learners with a suitable amount of interest, sympathy, or attention during their interaction.

System Overview

The current version of the system's front-end is implemented with the Unity game engine, and the back-end is implemented with Node.js. It can run on Mac OS and

Table 1 Implemented CSs in DiMaCA

Strategy	Description	Example
Simplification or Approximation	Use an alternative or a shorter term, which expresses the meaning of the target lexical item	Agent: May I have your order please? Learner: ... (silent) Agent: Order please.
Code-switching	Use of first language (L1) words with L1 pronunciation in second language (L2) conversation	Agent: May I have your order please? Learner: ... (silent) Agent: Go chimon wa ikaga desu ka? (Code-switching by using Japanese (learner's L1) words in English (learner's L2) conversation)
Asking clarification	Request an explanation of an unfamiliar meaning structure	Learner: One xxx please. Agent: What do you mean?
Suggesting AP (Answer Pattern)	Provide an example of answer that could fit the current discourse context	Agent: What would you like to drink? Learner: ... (silent) Agent: For example, you may say, 'one beer please' to order a beer.
Asking confirmation	Request confirmation that one heard or understood something properly	Agent: Check please! Learner: Do you mean the bill?
Asking repetition	Request repetition when one did not hear or understand something properly	Learner: May I have xxx? Agent: Would you please repeat that?
Expressing NU (Non-Understanding)	Express that one did not understand something properly	Learner: May I have xxx? Agent: I don't understand.
Guessing ^a	Assume the learner's answer among the expected answers in a current dialogue context based on the pre-defined dialogue agenda	Agent: Would you like anything else? Learner: The food is good. Agent: All right, call me anytime then. (Considering the learner is not willing to ask anything else because his answer does not correspond to Yes/No, nor to Food/Drink)
Repetition	Repeat the original utterance	Agent: How was your dinner? Learner: Pardon? Agent: How was your dinner?

^a An approximation of the original Guessing strategy was implemented in the current study. The original guessing strategy as described in Dörnyei and Scott's taxonomy (1997) implies for the speaker, when facing real indecision, to imagine what his interlocutor might be referring to from a wide range of possibilities. In our study, the agent does not actually perform itself a dynamic and exhaustive reasoning about the learner's intention from a wide range of possibilities, but rather performs it according to a narrow range of expected answers in the current conversation turn, as specified in advance in the dialogue script

Table 2 Implemented ABs in DiMaCA

Strategy	Description	Example
Congratulatory AB	Employed when the conversation with the learner is going well, as expected	<ul style="list-style-type: none"> • Okay, that's nice! • Great! • Ah, I see! • ...
Encouraging AB	Employed when the learner seems to hesitate to the extent that he/she remains silent	<ul style="list-style-type: none"> • Come on, you can do it! • You're almost there! • Don't be shy. • ...
Sympathetic AB	Employed when the learner's utterance does not match the agent's expectations.	<ul style="list-style-type: none"> • Sorry I couldn't get you dear. • Oops, I'm afraid I missed something. • ...
Reassuring AB	Employed when the learner seems to face much difficulties in the conversation	<ul style="list-style-type: none"> • Don't worry dear. • You'll be fine. • Let me help you. • ...

Windows. Figure 1 shows a screenshot of the system's interface, featuring the agent in a restaurant context.

The core architecture of the system was developed and implemented in our previous work (Ayedoun et al. 2016). It has two main components: the dialogue manager and multimodal response generator. Both are connected to several external web services and resources, as shown in Fig. 2 (top). Speech recognition is, for instance, performed by an external speech recognition service. We have, thus far, used Wit.ai¹ and Dialogflow,² both of which provide features to retrain the speech recognizer automatically, to obtain good quality in terms of speech recognition, even for L2 speakers. Both services can turn a learner's utterance into structured data, providing a semantic interpretation of the utterance in the form of an "intent" and "entities" (i.e., parameter values of the "intent") with a confidence score. Because the same "intent" (e.g., *Greeting*) can be expressed in multiple ways (e.g., *Hey*, *Good morning*, *Hello*), being able to directly grasp the learner's "intent" gives our system flexibility to handle various utterances that learners use to convey the same meaning. Of course, all the "intents" required for the conversation domain should be defined beforehand on the platform provided by Wit.ai or Dialogflow, so that they can be identified later from the learner's utterances. In our ECA, the detected "intent" from a given learner's utterance is considered to be the learner's intended meaning.

The overall conversational routine is supervised by the dialogue manager, which controls the various phases of dialogue and their timing, plus the level of system initiative, in an integrated fashion. The dialogue manager is composed of two modules, as shown in Fig. 2(top):

the dialogue flow management module (DFMM): which was implemented in our previous work (Ayedoun et al. 2016) and is responsible for controlling the

¹ Wit.ai (Facebook, Inc.): <https://wit.ai/> (accessed 2018/06/28)

² Dialogflow (Google, Inc.): <https://dialogflow.com/> (accessed 2018/06/28)



Fig. 1 Current version of the system interface featuring a conversational agent (Peter) in a restaurant context

dialogue flow (i.e., content related to the conversation task), based on the “intent” detected from the learner’s utterances and the dialogue script;

the strategies management module (SMM): which is a newly implemented module, whose mission is to trigger adapted CS and AB per the different states specified in DiMaCA. Hence, DiMaCA is the model specifying the set of rules implemented as the SMM module to make decisions about which conversational strategies to employ in the dialogue according to the conversation state (see Fig. 2 (bottom)). In the next section, we detail how the SMM triggers CS and AB adaptively and enhance the overall dialogue management ability of our conversational agent.

Conversational Strategies Enhanced Dialogue Management

Our system aims to make it easier for L2 learners to overcome their communicative difficulties and to make them feel at ease during conversation. However, the required conversational strategy depends upon the affective and behavioral state of the learners. If a learner faces a specific pitfall state, the system should diagnose the state and help the learner move into a state more conducive to conversation. When learners are in such a state, it is more likely that they will feel more motivated. Thus, the system needs to ensure that it is maintained. There is an abundance of literature on modeling affect and motivation (Afzal and Robinson 2011; Burleson and Picard 2007; D’Mello and Graesser 2012). It has been argued that, whereas a precise estimation of a learner’s specific state might not be possible or even required, an approximation of the state can be helpful in the sense that it can allow a system to be more empathetic, leading to higher levels of engagement (Suleman et al. 2016). Thus, our system does not aim to determine learners’ states accurately, but rather to approximate such states.

As described in Fig. 2 (bottom), the SMM (i.e., DiMaCA)’s routine goes from *Start* to *End* by checking the following states: *The learner is silent*; *The learner is asking for help*; *The learner is able to Understand but Not Answer (The learner is UNA)*; *The learner is Not able to Understand, Nor to Answer (The learner is NUNA)*; *The agent is*

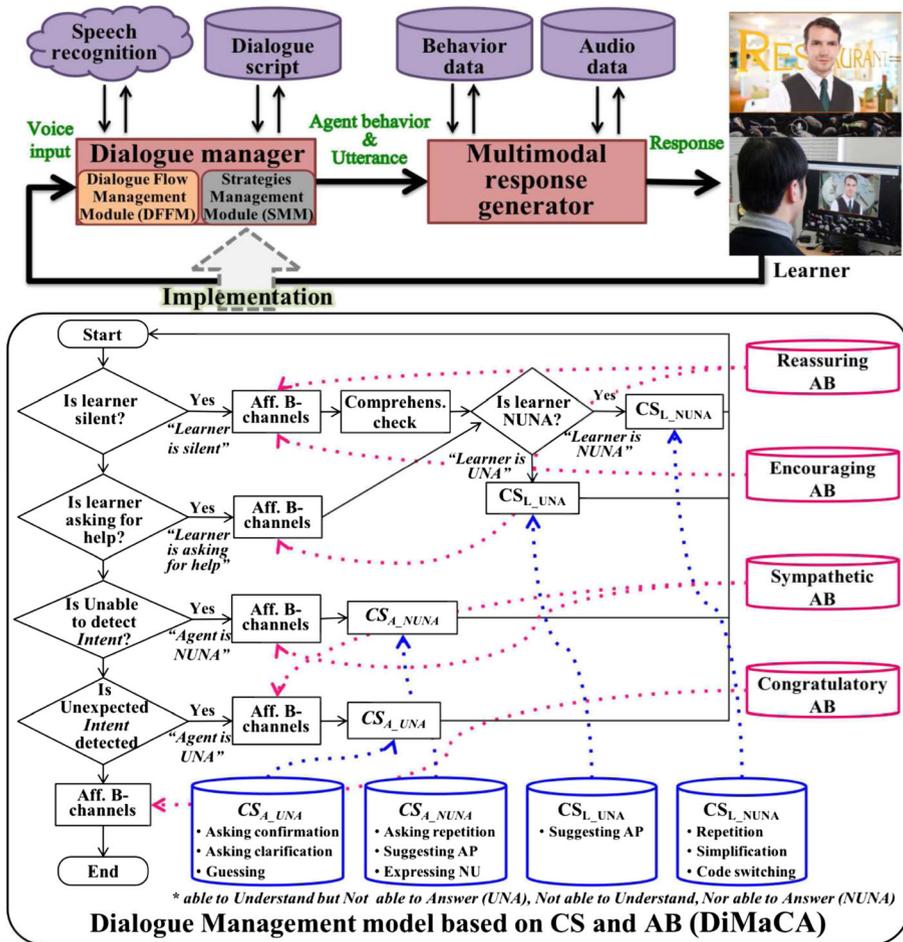


Fig. 2 System architecture showing interface (top) and DiMaCa (bottom)

able to Understand but Not Answer (The agent is UNA); and finally The agent is Not able to Understand, Nor to Answer (The agent is NUNA).

In our system, detection of these states automatically triggers adapted conversational strategies (indicated by the square symbols) that are retrieved from their respective database (indicated by the dotted lines) to keep the learner motivated by using AB (represented as database symbols in pink) and also to move the dialogue forward by using CS (represented as database symbols in blue).

When a given category of AB is triggered, the corresponding AB is chosen from the available options in that specific category in a stochastic way. On the other hand, an appropriate CS is chosen in a heuristically predefined order (e.g., top-to-bottom in the list of each CS category; Fig. 2 (bottom)). For example, when the conversation state, *The learner is NUNA*, is detected, the system first makes use of *Repetition*. If the same state is detected on the next turn, then *Simplification* is applied. If the same state is detected on the following turn, *Code-switching* is used. The reason for the order is to make it progressively easier for the learner to overcome their current difficulty when the

conversation is stuck in each state. The estimation of each state, and, consequently, the firing of specific conversational strategies, is done as follows.

The learner is silent: occurs when the system is expecting input from the learner, but has not gotten anything after a certain period has elapsed. In this case, the system first applies a *Reassuring* or *Encouraging* AB and then tries to determine whether the learner is NUNA or UNA. To that end, the system investigates why the learner is silent by conducting a comprehension check using simple questions such as “*Do you understand?*” Upon interpreting the learner’s answer as positive or negative, the system concludes either that the learner is silent because he or she does not understand what is being requested or that he or she does not know how to answer. Moreover, if the learner remains silent, even after performing a comprehension check, the system randomly assumes that the learner is either NUNA or UNA and fires per the respective CS. This enables the system to always actively try new strategies to help learners overcome their communication pitfalls or fears.

The learner is NUNA: occurs when the learner is unable to understand what the agent expects from him/her. This state is determined when the learner gives a negative answer after the system performs a comprehension check or by the system detecting such an intended meaning from the learner’s utterance (e.g., “*I don’t understand*” “*Please repeat*” or “*Pardon?*”). In such case, the system will fire a specific CS_{L_NUNA} , such as *Repetition*, *Simplification*, or *Code-switching*, per the selection policy described earlier.

The learner is UNA: occurs when the learner understands what is being requested of him/her, but cannot or does not know how to answer. This state is determined when the system gets a positive answer from the learner after it has performed a comprehension check or by it detecting such an intended meaning from the learner’s utterance (e.g., “*I don’t know*”). When *The learner is UNA*, the agent fires the CS_{L_UNA} *Suggesting AP* to give the learner a hint about how to overcome his/her current difficulty.

The learner is asking for help: occurs when the learner expresses that he or she is NUNA, UNA, or specifically requests a CS, such as *Repetition* or *Simplification*. In this case, the system will fire a *reassuring* AB and then apply the appropriate CS per the nature of the help requested by the learner.

The agent is NUNA: occurs when the system is unable to detect the learner’s intended meaning because of a low confidence score or when a recognition error of the learner’s utterance occurs. In this case, the system will first fire a *Sympathetic* AB and then try to recover by applying a specific CS_{A_NUNA} , such as *Asking repetition*, *Suggesting AP*, or *Expressing NU*, to give the learner another chance to express himself or herself. The appropriate CS_{A_NUNA} is chosen according to the selection policy described earlier.

The agent is UNA: occurs when the system detects the learner’s intended meaning with an acceptable confidence score, but is unable to give an answer to him/her in the current context. For example, the learner asks for the nearest supermarket, but the agent expects him/her to place an order in a restaurant context. In this case, the agent will first apply a *Sympathetic* AB and then try to reformulate the intended meaning by using a specific CS_{A_UNA} such as *Asking for confirmation* or *Asking for clarification*, to make sure that what it has understood from the learner’s

utterance is what he or she meant or *Guessing* to make the conversation moves forward in the last resort, per the selection policy.

The SMM is triggered each time an utterance is detected from the learner or when a certain amount of time has elapsed after the system requested an answer from the learner (i.e., *the learner is silent*). When one of the six conversation states mentioned earlier is detected, the corresponding CS and AB are triggered. As described above, the following cues are used for estimating the conversation state: whether the learner is silent; whether an intent has been detected from his/her utterance; whether the confidence score of the detected intent is high; whether the learner is explicitly saying he or she is NUNA or UNA; and whether the system can provide an answer to the detected intent. In case none of the conversation states (i.e., *The learner is silent*; *The learner is asking for help*; *The learner is able to Understand but Not Answer (The learner is UNA)*; *The learner is Not able to Understand, Nor to Answer (The learner is NUNA)*; *The agent is able to Understand but Not Answer (The agent is UNA)*; and finally *The agent is Not able to Understand, Nor to Answer (The agent is NUNA)*) is detected, we deduce that 1) the learner has uttered something and was not asking for help (since the learner is neither NUNA nor UNA) and that 2) the agent understood the learner's utterance and can answer it (since the agent is neither NUNA nor UNA). Hence, from 1) and 2), the system concludes that a successful turn has occurred, performs a *Congratulatory AB*, and the lead for dialogue management is shifted to the DFMM. The DFMM determines the system's next utterance on the basis of content detected from the learner's utterance (i.e., *intent, entities*) and the dialogue script. This routine is repeated in a loop until the end of the conversation.

To achieve dialogue management, the dialogue manager keeps information in memory, as follows:

The conversation's current step: is used by the DFMM to determine the current stage of the conversation and to carry out the system's next move using the specifications of the dialogue script such as next utterance to fire, specific scene events (e.g.: displaying the menu, moving to another area, etc., that are requested to be performed in the scene), and waiting time. Moreover, the dialogue script is written in XML and can be easily modified or replaced to target a different conversation context.

The system's previous utterance: is used to perform CSs such as *Repetition*, *Simplification*, or *Code-switching*, when requested by the SMM.

Triggered CS and AB on the current conversation state: is used to determine a CS or AB that has not yet been tried, but could be used if the current state is successively detected several turns in a row.

We expect that the modular and domain-independent nature of the proposed dialogue management model (i.e., DiMaCA) will not only facilitate its reusability across different dialogue domains, but it will also make it easier to develop conversational spoken language interfaces that are more adaptable to L2 learners from the WTC standpoint.

Figure 3 shows an excerpt of a conversation between a learner and the agent depicting how AB (pink) and CS (blue) are called into action, per the different dialogue states (gray). As shown, the successive interventions of the system are successful in

Agent: Good evening. Do you have a reservation?
 Learner: No.
 Agent: **Ah, I see!** [AB: *Congratulatory*]
 Would you prefer smoking or non-smoking table?
 Learner: ... (“Learner is silent”)
 Agent: **Come on, don’t give up!** [AB: *Encouraging*]
 Do you understand? [CS: *Comprehension check*]
 Learner: No. (“Learner is NUNA”)
 Agent: **Which area do you prefer?** [CS_{L_NUNA}: *Simplification*]
 Learner: I don’t know area. (“Learner is UNA”)
 Agent: **Don’t worry, I’m here to help you.** [AB: *Reassuring*]
 For example, you could say: “Non-smoking area please!”
 [CS_{L_UNA}: *Suggesting AP*]
 Learner: xxx xxx please! (“Agent is NUNA”)
 Agent: **Sorry, I’m afraid I missed something.** [AB: *Sympathetic*]
 Could you please repeat that? [CS_{A_NUNA}: *Asking repetition*]
 Learner: Non-smoking please.
 Agent: **Great!** [AB: *Congratulatory*]
 Please come this way...

Fig. 3 Excerpt of actual dialogue between the agent and a learner illustrating usage of AB and CS

helping the learner gradually overcome his initial breakdown, following the question “would you prefer a smoking or non-smoking table?” As illustrated, it is the successful estimations of the occurring pitfalls, combined with the help provided by the agent through usage of appropriate AB and CS that finally leads to a conceivable answer, “non-smoking please,” from the learner. Without such strategies, the conversation would probably end just after the agent’s first question, because the learner seems unable to proceed with the interaction. It is this kind of support that our dialogue agent aims to provide to L2 learners through the use of ABs and CS.

Pilot Study

We conducted a pilot study to answer the following research questions:

- 1) “Which combination or single implementation of CS and AB have the potential to enhance L2 learners’ willingness to communicate?”
- 2) “Are there similar tendencies between 1) (i.e., WTC tests results) and learners’ opinion on the support provided by the system through CS and or AB?”

Experimental Design

To answer these two questions, the flow of the experiments was designed according to two phases so as to compare learners’ WTC results across different versions of the

system on one hand (Phase 1), and examine their preference after interacting with different versions of the system, on the other (Phase 2), as shown in Table 3.

During Phase 1, we gauged learners’ WTC by administering a self-report survey before and after their first interaction with the system. Multiple system versions, including the normal version featuring both CS and AB (CS + AB), a second version featuring only CS, and version third featuring only AB, were employed in the interactions so that we could examine how the outcomes on participants’ WTC varied with the system version.

To complete Phase 2, we let all participants have a second round of interactions with another version (i.e., different from the one used for their first interaction) of the system. We then conducted a survey to get feedback concerning their preference on the system’s versions and thereby examine whether there would be similar tendencies between the WTC test results and what learners actually felt about the support provided by the system.

To minimize the eventuality that learners’ preference would be due to only the order in which they interacted with different versions of the system (i.e., order effect), the learners’ interactions with the system in each group were designed by applying the counterbalancing method proposed by Howitt and Cramer (2011). To this extent, participants of Group 1 were split into two groups of ten participants for the second round of their interaction with the system; half of them interacted with the CS version, while the other half interacted with the AB version, as shown in Step 4 of Table 3.

Instruments and Steps

Conversational Agent

We used an embodied conversational agent that was implemented on the system proposed in our previous work (Ayedoun et al. 2016) and enhanced it with the management model (i.e., DiMaCA) described above. The system makes possible spoken dialogues between the conversational agent, personified as Jack, and learners in a restaurant context, as illustrated in Fig. 4. The conversation scenario begins with an entrance scene where learners are welcomed by Jack. After checking whether they have a reservation or not, they are guided to a table in their preferred area (smoking, non-smoking). From there, learners can call Jack anytime, ask for the menu, order drinks,

Table 3 Overview of the experiment flow

Phase	Steps	Group 1(G1)	Group 2(G2)	Group 3(G3)	
Phase 1	Step 0	First WTC questionnaire (Pretest)			
	Step 1	Warm-up interaction with the system			
Phase 2	Step 2	CS + AB(n = 20)	CS (n = 10)	AB(n = 10)	
	Step 3	Second WTC questionnaire (Posttest)			
	Step 4	<i>G1a</i> CS(n = 10)	<i>G1b</i> AB(n = 10)	<i>G2</i> CS + AB (n = 10)	<i>G3</i> CS + AB(n = 10)
	Step 5	System preference survey			

dishes of their preference, and request the bill, just as they would do in a restaurant. During the interaction, learners were able to answer Jack's questions or take the initiative to ask questions or make orders. Note that Jack could also make use of common non-verbal backchannels like gaze, head nods, smiles, and blinking to some extent. Please refer to Ayedoun et al. (2016) for a description of the agent's non-verbal backchannel cues.

Participants

The study was conducted with 40 (24 males and 16 females) Japanese undergraduates and graduate students attending a Japanese university. In terms of language background, the participants were quite homogeneous; all of them were native Japanese speakers and none had lived in an English-speaking country. They were informed that their participation in the study was voluntary and that the results would be anonymized.

The evaluation was conducted in six steps, as listed in Table 3.

Step 0 (First measure of WTC, Pretest): we employed a widely used survey (i.e., Cronbach $\alpha = .88$) developed by Matsuoka (2006) and inspired by Sick and Nagasaka's WTC test (2000), to gauge learners' WTC before (Procedure 1) their interaction with the system. The WTC survey targeted three variables: *confidence*, *anxiety*, and *desire to communicate*, which are considered to be the immediate precursors of WTC (MacIntyre and Charos 1996). Participants were asked to rate 30 scenarios (e.g., *making a telephone call to make a reservation at a hotel in an*



Fig. 4 Learners interacting with the conversational agent

English-speaking country) related to using English in various circumstances on a four-point Likert scale (0–3). An exhaustive list of the 30 situations is presented in the English version (Ockert 2012) of the WTC survey in the Appendix Table 9. As described in Table 4, the first variable tests for *confidence* and asks participants to rate the scenarios from 0 (*I couldn't do that*) to 3 (*I could do that easily*). The second variable tests for *anxiety* and asks participants to rate the same scenarios from 0 (*I would definitely not be anxious*) to 3 (*I would be extremely anxious*). Finally, participants were asked to rate the third variable, *desire to communicate* in English from 0 (*I wouldn't want to try that*) to 3 (*I would absolutely want to try that*). All participants were given as much time as needed to complete the questionnaires. Data were collected anonymously via an online survey service, and participants were told that their answers would be kept confidential. The scores for *confidence*, *anxiety*, and *desire* were derived by calculating the average of the sum of learners' self-reported ratings for each subscale. Then, participants were split into three groups (Group 1 to 3): the most uniform possible in terms of *confidence*, *anxiety*, and *desire to communicate*.

Step 1 (Warm-up interaction with the system): All participants were initially asked to interact with Jack (the agent), who would teach them how to pronounce some words in English. They were requested to listen and repeat the words per Jack's instructions. Our intention was to let all the learners get acquainted with the agent and understand how the system works.

Step 2 (First interaction with the system): Participants in each group were then asked to interact with the system. The conversation was held in a restaurant context with Jack interacting with them as a waiter. The three different versions of the system: CS + AB, CS only, and AB only. Participants interacted with a given version of the system per their group. For example, participants in Group 1 interacted with the CS + AB version, while those in Group 2 with the CS version, as indicated in Table 3. Note that participants interacted individually with the system in a room specially prepared for the evaluation and were given as much time as they wished to enjoy the conversation with Jack, until the end of the interaction. They were also informed that they were free to interrupt the interaction at any time they desired, but were requested to let us know beforehand.

Step 3 (Second measure of WTC, Posttest): The WTC survey used after the first interaction with the system was similar to the survey described in Step 1, but differed in that participants were asked to rate the same scenarios, imagining their

Table 4 Rating scale used for first WTC survey (Pretest)

Scale	confidence	anxiety	desire to communicate
3	I could do that easily.	I would be extremely anxious.	I would absolutely want to try that.
2	I could do that.	I would be anxious.	I would want to give it a try if given the chance.
1	I don't know if I could do that.	I would be a little anxious.	I don't know if I would want to try that.
0	I couldn't do that.	I would definitely not be anxious.	I wouldn't want to try that.

level of *confidence*, *anxiety*, and *desire to communicate* if given the opportunity to frequently converse with our conversational agent. Table 5 shows the rating scales used in this survey. In the first section of the test, learners were asked to evaluate their expected *confidence* from 0 (*I don't think I'd be able to do that*) to 3 (*I think I'd be able to do that easily*). In the second section, they were asked to rate their expected *anxiety* from 0 (*I don't think I'd be anxious*) to 3 (*I think I'd be extremely anxious*). In the third section, they rated their expected *desire to communicate* in English from 0 (*I don't think I'd feel like trying*) to 3 (*I think I'd absolutely be willing to try that*).

Step 4 (Second interaction with the system): After taking the second WTC questionnaire (posttest) in Step 3, participants were instructed to interact again with the system in the same restaurant context. As in Procedure 2, participants interacted with different versions of the system according to their group but were not informed that the system was different from the one they used in their first interaction. Participants in Group 1 were randomly split into two groups (G1a and G1b) of 10 participants each and assigned a different system version (CS for G1a and AB for G1b participants), as shown in Table 3.

Step 5 (System preference survey): Following Procedure 4, all participants were asked to choose which of their two interactions (i.e., which version of the system) they preferred the most and the reason for their choice.

Results

Phase 1 (WTC Measures) Results:

We first conducted a Levene's test to assess the homogeneity of variance of WTC results among the three groups. The results revealed that the null hypothesis (i.e., equal variance among the three groups) stands ($p > .05$ for the three variables (i.e., *confidence*, *anxiety*, *desire to communicate*)), and therefore the homogeneity of variances among the three groups was not violated. A one-way ANCOVA was then conducted to determine whether there were statistically significant differences between the WTC posttest results of each group in terms of expected *confidence*, *anxiety*, and *desire to communicate* whilst controlling for pretest results. Post-hoc Tukey Kramer tests were additionally run to further investigate the differences.

Table 5 Rating scale used for second WTC survey (Posttest)

Scale	confidence	anxiety	desire to communicate
3	I think I'd be able to do that easily.	I think I'd be extremely anxious.	I think I'd absolutely be willing to try that.
2	I think maybe I'd be able to do that.	I think I'd be anxious.	I think maybe I'd be willing to try.
1	I don't know if I'd be able to do that.	I think I'd be a little anxious.	I don't know if I'd feel like trying.
0	I don't think I'd be able to do that.	I don't think I'd be anxious.	I think I'd not feel like trying.

There was a significant difference in participants' expected *confidence* [$F(2, 36) = 7.139, p < .05$] among the three groups. The post-hoc Tukey Kramer tests (Table 6) showed that Group 1's *confidence* was significantly higher than that of Group 2 ($p = 0.002$). Group 3's *confidence* tended to be significantly higher than that of Group 2 ($p = 0.077$).

There was a trend towards significant difference in participants' expected *anxiety* [$F(2, 36) = 3.472, p < .1$] among the three groups. The post hoc tests showed (Table 7) that Group 1's *anxiety* tended to be significantly lower than Group 2's ($p = 0.072$).

There was a significant difference in participants' expected *desire to communicate* [$F(2, 36) = 6.466, p < .05$] among the three groups. The post hoc tests (Table 8) showed that Group 1's *desire to communicate* was significantly higher than Group 2's ($p = 0.012$) and Group 3's ($p = 0.011$).

In total, as shown in Fig. 5, the between-groups analysis revealed that Group 1's (CS + AB) results were significantly better than Group 2's (CS) in terms of expected *confidence*, and *desire to communicate*; Group 1's results tended to be better than those of Group 2 in terms of expected *anxiety*. Moreover, Group 1's results were also significantly better than Group 3's (AB) in terms of *desire to communicate*. Finally, Group 3's results tended to be better than those of Group 2 in terms of *confidence*.

In addition, a one-way ANOVA was conducted to compare the effect of the system version on the durations of the participants' interactions. Even though no time constraints were given, we could not find any significant differences between groups regarding the amount of time that their participants spent on the task in the experiments [$F(2, 37) = 0.28, p = 0.75$].

Phase 2 (System Preference Survey) Results

The preference rate of the **CS + AB version was uniformly high** across all four groups (note that Group 1's participants were split in two groups, i.e. G1a and G1b, as mentioned earlier); this version was preferred by 32 participants out of 40 (80%) in total, whereas the CS and AB versions were preferred respectively by five participants out of 20 (25%) and by three participants out of 20 (15%), as shown in Fig. 6. Note that this tendency was observed across all four groups, no matter the order in which learners interacted with the CS + AB version. In fact, as reasons justifying their choices, participants who preferred the CS + AB version frequently mentioned that they found it natural and warm the way Jack showed empathy throughout the interactions and appreciated the help they got from him when facing difficulties in understanding or expressing what they had to say.

Table 6 Pairwise comparisons of *confidence* scores across all groups

Group pairs	Adj. mean	Std. error	Q	p-value	Cohen's d
G1-G2	0.46	0.09	5.20	0.002	1.42
G1-G3	0.14	0.09	1.54	0.527	0.42
G3-G2	0.32	0.10	3.17	0.077	1.00

Table 7 Pairwise comparisons of *anxiety* scores across all groups

Group pairs	Adj. mean	Std. error	Q	p-value	Cohen's d
G1-G2	0.26	0.08	3.22	0.072	0.88
G1-G3	0.19	0.08	2.38	0.225	0.65
G3-G2	0.07	0.09	0.73	0.864	0.23

Discussion

The above results allow us to draw a number of preliminary conclusions. First, the combination of CS and AB proved to be promising in motivating L2 learners towards communication in the target language, much more than a version of the system containing either CS or AB. This is corroborated by the results of the self-reported WTC measures analysis and those of the preference survey, confirming our initial beliefs that making possible smooth and interactive conversations is not, by itself, sufficient to increase L2 learners' WTC. This also requires the ability to convey enough warmth or sympathy to learners during interactions. The proposed dialogue management model in this paper (i.e., DiMaCA) covered both requirements by the way of CS and AB, and the results obtained are meaningful in terms of validating our approach. Also, in our previous work (Ayedoun et al. 2016), a version of the system that contained the conversational agent without DiMaCA led to lower expected WTC than in the current study, which supports the view of the potential usefulness of the model proposed in this paper. Thus, we feel that the results enable us to tentatively conclude that the participants in this study would display actual WTC gains if given opportunities to interact frequently with this kind of system. Since WTC is believed to have an immediate and sustained influence on learners' actual usage frequency of the target language (MacIntyre et al. 1998), it is important to create environments dedicated to enhancing L2 learners' WTC. It seems that, for the participants in this study, the environment offered by the conversational agent could help to enhance their willingness to communicate in the target language.

Surprisingly, these results also suggest that even the version of the system that only contained AB strategies, which was introduced to alleviate expected anxiety, was potentially more effective at enhancing L2 learners' expected confidence than a version of the system that only contained CS strategies. This is an interesting finding, because it seems to confirm prior observations by MacIntyre and his colleagues (MacIntyre and Charos 1996), who reported a positive correlation between a decrease in anxiety and increase in confidence among L2 learners. This might also reveal that the empathetic

Table 8 Pairwise comparisons of *desire to communicate* scores across all groups

Group pairs	Adj. mean	Std. error	Q	p-value	Cohen's d
G1-G2	0.55	0.13	4.29	0.012	1.17
G1-G3	0.56	0.13	4.34	0.011	1.18
G3-G2	0.01	0.15	0.04	0.999	0.01

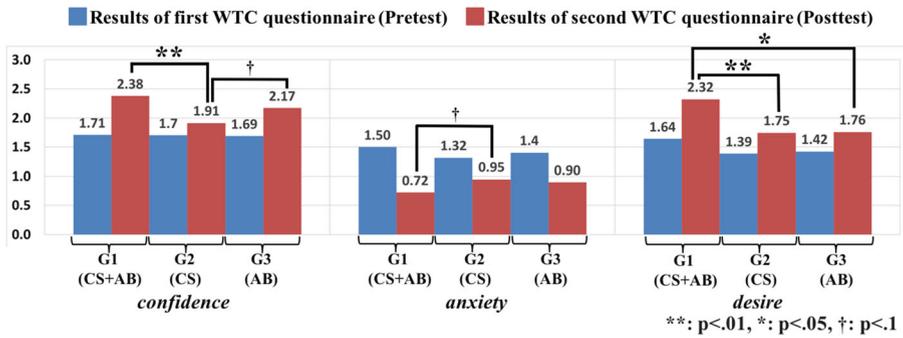


Fig. 5 Between-groups analysis results

support that L2 learners receive during interactions might be an important factor influencing both their anxiety and confidence, and consequently their willingness to communicate in the target language. Thus, a conversational agent providing a careful empathetic support to learners by reassuring, encouraging, and congratulating them when necessary, might be effective in enhancing their willingness to communicate in the target language.

In addition, even though no time constraints were placed on the participants, we could not find any significant differences between groups regarding the amount of time they spent on task in the experiments, contrary to the results for expected WTC which differed from one group to another. This is also an interesting outcome that might actually help us further understand the role played by additional factors such as duration of interaction in promoting WTC-friendly conversations between L2 learners and a conversational agent.

The discussion raises the question of whether employing conversational strategies embedded in ECAs should be promoted as part of the second language learning process, or even integrated into the curriculum and whether features of ECAs should be used in classroom teaching. Although our study does not directly attempt to address these questions, we do feel that our work contributes to the body of research in the sense that it gives hints on the potential of ECAs embedding specific conversational strategies to facilitate essential elements of the second language acquisition process and, as such, deserves more attention.

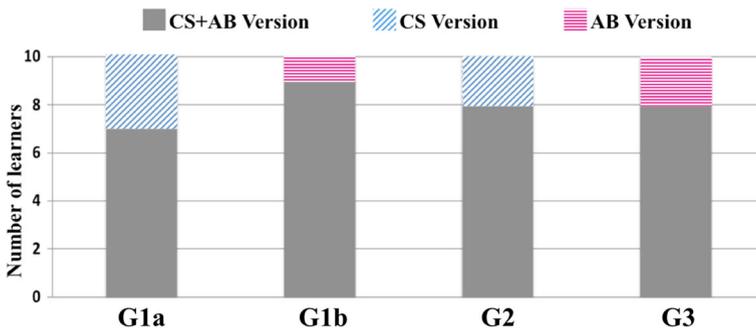


Fig. 6 System preference survey results

Limitations and Future Work

Although this study has reached its aims, it has some limitations.

First, since L2 learners' WTC does not, of course, increase overnight, our study at its current stage could not measure actual WTC gains among L2 learners. It was rather dedicated to collecting insights about the potential of using an embodied conversational agent coupled with specific conversational strategies towards increasing L2 learners' WTC. We should bear in mind that the WTC questionnaires (i.e., Pretest and Posttest), although asking similar questions, were different in the sense that the first asked for learners' actual WTC, whereas the second asked about learners' expected WTC after using the system for a while. Moreover, the novelty factor may also have affected the results of our experiment. Although we tried to minimize this by familiarizing learners with the conversational agent before their first interaction, the fact that being able to simulate a daily conversation with an ECA is unusual might have added a degree of excitement and might have colored participants' answers. Hence, in the future, we plan to carry out long-term evaluations to collect more reliable data on actual WTC gains and examine the relationship between improvement of WTC and outcomes of our approach on learners' actual involvement in L2 communication, since the WTC measures used in the current study were strictly self-reported. We will also need to increase the sample size of our experiments and design the different experiments conditions so as to also have groups of learners interacting exclusively with CS and AB versions of the system. An analysis of such learners' system preference might help us deepen our understanding of L2 learners' actual perception of the support provided by each strategy (i.e., CS and AB).

Another possible limitation of this study is that the current version of our system is limited to conversations in only one context (e.g., a restaurant context). We are currently developing an authoring tool that will make it easier to design and implement new conversation scenarios in order to offer opportunities for learners to converse in various contexts.

Furthermore, in this paper, we did not actually report data on the participants' interactions, such as the number of CS or AB fired in each condition. We acknowledge that such data might be useful to the extent to help us clarify the practical impact of such strategies on L2 learners' WTC. Nevertheless, in order to conduct a careful discussion on this point, we consider that we should design a comprehensive analysis methodology that goes beyond a simple comparison of learners' conversations log data but also take into account monitoring data of their internal states during the conversations. In the same vein, the excessive use of AB strategies might be perceived as "heavy handed", while infrequent use of AB strategies may not provide L2 learners with the support and encouragement they need in order to attain a higher WTC. Besides, given that ABs are randomly chosen, inappropriate AB responses may occur depending on the context. Therefore, we should also take a close look at the rate of appropriate/inappropriate AB as well as CS responses by conducting in-depth analysis of interaction log data and examining fluctuation in learners' reactions to the system on a turn-by-turn basis. Additionally, we should investigate alternate ways of timing the relative frequency of such AB and CS strategies. Ultimately, we would like our system to converse by dynamically taking into consideration previous interactions and learners' L2 competence, and by deploying conversational strategies in accordance with their affective state and level of WTC. To that extent, we plan to use non-verbal cues such as learners' facial expressions for estimating their affective state.

Conclusion

Most intelligent tutoring systems and particularly spoken dialogue interfaces dedicated to L2 learning do not explicitly consider aspects related to affective variables influencing learners' engagement towards the target language production. This paper described DiMaCA, a dialogue management model based on two conversational strategies (i.e., communication strategies (CS) and affective backchannels (AB)) aiming to empower embodied conversational agents to foster L2 learners' willingness to communicate in an English-as-a-foreign-language context.

The pilot study results showed that the combination of CS and AB could be effective, because participants who interacted with the CS + AB version of the system displayed higher expected WTC than those who interacted with the CS only or AB only versions. We also found that even the AB-only version of the system had the potential to enhance L2 learners' expected WTC to a certain extent, suggesting that the empathic care an ECA provides to L2 learners might be essential in raising their WTC. Future research should be directed to confirming the tendencies mentioned above by evaluating in more detail the effects associated with each strategy (i.e., CS or AB), determining approaches for strengthening their impact on L2 learners' WTC, and carrying out necessary long-term evaluations of their outcomes on learners' actual involvement in L2 communication, since the WTC measures used in the current study were strictly self-reported.

For countries like Japan, where English learning focuses less on development of communicative skills and where learners have limited access to opportunities for using the target language, the results obtained in our study provide insights on the meaningfulness of integrating embodied conversational agents in the traditional teaching curriculum. We hope that this work will have genuine value and contribute to the development of more effective computer-based intelligent approaches to enhance WTC among L2 learners.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix

Table 9 English Translation of the conversation situations in Matsuoka (2006)'s WTC Survey (Ockert 2012)

- 1 Asking a Japanese teacher for a copy of an audio recording.
- 2 Complaining to a Japanese teacher that the speed of the listening test was too quick to catch.
- 3 Complaining to a native teacher that the speed of the listening test was too quick to catch.
- 4 Giving a reply for an American television program covering student life in Japan.
- 5 Making a telephone call in order to make a reservation at a hotel in English speaking country.
- 6 Interviewing a native English speaker for an article in the school paper.
- 7 Asking a pair work partner for the time now.
- 8 Speaking to a foreigner sitting next to you on the train.

Table 9 (continued)

9	Asking a native English speaking teacher the meaning of a word.
10	Making a phone call to invite a friend who can speak only English to a party.
11	Asking a native teacher for a handout given when you were absent from class.
12	Talking to your pair work partner about a TV program which you watched.
13	Stand in front of the entire class and talk about a TV program which you watched.
14	Helping a foreigner that looks troubled because he cannot read a restaurant menu.
15	Asking a foreigner for the time when you do not know it.
16	Help a troubled foreigner because he cannot understand what the salesclerk says at the supermarket
17	Greet a group of medical professionals who came from the United States to visit your school.
18	In front of your class, answer a native teacher's questions about your trip during summer vacation.
19	Stand in front of your class and talk about your memories of your summer vacation for two minutes.
20	To buy a rare CD sold only overseas, call a CD store in the United States by telephone to order one.
21	Take a small number of English speaking people sightseeing in Tokyo for one day.
22	Call your host family and thank them for letting you stay with them.
23	Tell your pair work partner in English the way to a place using a map.
24	Say five English words which start with S to your pair work partner.
25	Ask a native English-speaking teacher to copy a CD.
26	Ask the meaning of a word to a Japanese teacher using classroom English.
27	Stand and tell your entire class five words using classroom English.
28	Talk to your pair work partner about your memories of summer vacation for two minutes.
29	Help a foreigner who looks troubled at the station.
30	Participate in an English language speech contest for Japanese students. Judges are native speakers.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Afzal, S., & Robinson, P. (2011). Natural affect data: Collection and annotation. In *New perspectives on affect and learning technologies*, 55–70. Springer New York.
- André, E., Rist, T., & Müller, J. (1998). Integrating reactive and scripted behaviors in a life-like presentation agent. In *Proceedings of the second international conference on Autonomous agents*, 261–268. ACM.
- Arslan-Ari, I. (2010). Animated pedagogical agents from students' perspectives. *Proceedings of the Annual Convention of the Association for Educational Communications and Technology*, 1, 6–9.
- Atkinson, R. K., Mayer, E. M., & Merrill, M. M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology*, 30(1), 117–139.
- Ayedoun, E., Hayashi, Y., & Seta, K. (2016). Web-services based conversational agent to encourage willingness to communicate in the EFL context. *Journal of Information and Systems in Education*, 15(1), 15–27.

- Bälter, O., Engwall, O., Öster, A. M., & Kjellström, H. (2005). Wizard-of-Oz test of ARTUR: a computer-based speech training system with articulation correction. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, 36–43. ACM.
- Banchs, R. E., & Li, H. (2012). Iris: a chat-oriented dialogue system based on the vector space model. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL) System Demonstrations*, 37–42.
- Bialystok, E. (1983). Some factors in the selection and implementation of communication strategies. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 100–118). New York: Longman Inc..
- Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(2), 293–327.
- Bohus, D., & Rudnicky, A. I. (2009). The ravenclaw dialog management framework: Architecture and systems. *Computer Speech & Language*, 23(3), 332–361.
- Burleson, W., & Picard, R. (2007). Evidence for gender specific approaches to the development of emotionally intelligent learning companions. *IEEE Intelligent Systems*, 22(4), 62–69.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47.
- Cassell, J., Bickmore, T., Campbell, L., Villhjálmsson, H., & Yan, H. (2000). Conversation as a system framework: Designing embodied conversational agents. *Embodied conversational agents*, 29–63.
- Chapelle, C. A. (2009). The relationship between second language acquisition theory and computer-assisted language learning. *The Modern Language Journal*, 93, 741–753.
- Clément, R., Baker, S. C., & MacIntyre, P. D. (2003). Willingness to communicate in a second language: The effects of context, norm, and Validity. *Journal of Language and Social Psychology*, 22(2), 190–209.
- Compton, L. (2004). Using text chat to improve willingness to communicate. In J-B Son (Ed.), *Computer-Assisted Language Learning: Concepts, Contexts and Practices*, 123–144, New York: iUniverse, Inc.
- D’Mello, S., & Graesser, A. (2012). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(4), 23.
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners’ fluency and comprehensibility development. *Applied Linguistics*, 29(3), 359–380.
- Dittmann, A. T., & Llewellyn, L. G. (1968). Relationship between vocalizations and head nods as listener responses. *Journal of Personality and Social Psychology*, 9(1), 79–84.
- Dörnyei, Z., & Scott, M. L. (1997). Communication strategies in a second language: Definitions and taxonomies. *Language Learning*, 47(1), 173–209.
- Dörnyei, Z., & Thurrell, S. (1991). Strategic competence and how to teach it. *ELT Journal*, 45(1), 16–23.
- Drolet, A. L., & Morris, M. W. (1998). Rapport in conflict resolution: Accounting for how nonverbal exchange fosters coordination on mutually beneficial settlements to mixed motive conflicts. Graduate School of Business, Stanford University.
- Duncan, S., & Fiske, D. W. (2015). Face-to-face interaction: Research, methods, and theory (Vol. 3). Routledge.
- Faerch, C., & Kasper, G. (1983). On identifying communication strategies in interlanguage production. *Strategies in interlanguage communication*, 210–238.
- Fraser, N. (1997). Assessment of interactive systems. In D. Gibbon, R. Moore, & R. Winski (Eds.), *Handbook of standards and resources for spoken language systems* (pp. 564–614). New York: Mouton de Gruyter.
- Fries, C. (1952). *The structure of English*. New York: Harcourt Brace.
- Goldberg, S. B. (2005). The secrets of successful mediators. *Negotiation Journal*, 21(3), 365–376.
- Hayashi, T., & Hayashi, R. (1991). Back channel or main channel: A cognitive approach based on floor and speech acts. *Pragmatics and language learning*, 2(1), 119–138.
- Howitt, D., & Cramer, D. (2011). Introduction to research methods in psychology. 3(164), 179–181. Harlow, Essex: Pearson Education Limited.
- Ito, K. (2000). Cooperative strategies by elementary and intermediate level Japanese language learners, bulletin of the international student center Gifu University, 38–52.
- Johnson, W. L., & Valente, A. (2008). Tactical Language and Culture Training Systems: Using Artificial Intelligence to Teach Foreign Languages and Cultures. In AAAI (pp. 1632–1639).
- Kang, S. J. (2005). Dynamic emergence of situational willingness to communicate in a second language. *System*, 33(2), 277–292.
- Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica*, 26(1), 22–63.

- Kopp, S., Allwood, J., Grammer, K., Ahlsen, E., & Stocksmeier, T. (2008). Modeling embodied feedback with virtual humans. In I. Wachsmuth & G. Knoblich (Eds.), *ZiF research group international workshop. LNCS (LNAI), 4930* (pp. 18–37). Heidelberg: Springer.
- Kopp, S., Gesellensetter, L., Krämer, N. C., & Wachsmuth, I. (2005). A conversational agent as museum guide—design and evaluation of a real-world application. In *International workshop on intelligent virtual agents* (pp. 329–343). Berlin, Heidelberg: Springer.
- Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: affective impact of animated pedagogical agents. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, 359–366. ACM.
- Liskin-Gasparro, J. E. (1996). Circumlocution, communication strategies, and the ACTFL proficiency guidelines: An analysis of student discourse. *Foreign Language Annals*, 29(3), 317–330.
- Long, M. H. (1981). Input, interaction, and second-language acquisition. In H. Winitz (Ed.), *Native language and foreign language acquisition* (pp. 259–278). New York: New York Academy of Sciences.
- MacIntyre, P., Baker, S., Clément, R., & Conrod, S. (2001). Willingness to communicate, social support, and language-learning orientations of immersion students. *Studies in Second Language Acquisition*, 23(3), 369–388.
- MacIntyre, P. D., & Charos, C. (1996). Personality, attitudes, and affect as predictors of second language communication. *Journal of Language and Social Psychology*, 15, 3–26.
- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *Modern Language Journal*, 82(4), 545–562.
- Matsuoka, R. (2006). Willingness to communicate in English among Japanese college students. *Proceedings of the 9th Conference of Pan-Pacific Association of Applied Linguistics*. 165–176.
- McCroskey, J. C. (1997). Willingness to communicate, communication apprehension, and self perceived communication competence: Conceptualizations and perspectives. In J. Daly et al. (Eds.), *Avoiding communication: Shyness, Reticence, & Communication Apprehension* (pp. 75–129). Cresskill: Hampton Press.
- Mesgarshahr, A., & Abdollahzadeh, E. (2014). The impact of teaching communication strategies on EFL learners' willingness to communicate. *SSLT*, 4(1), 51–76.
- Morency, L. P., de Kok, I., & Gratch, J. (2010). A probabilistic multimodal approach for predicting listener backchannels. *Autonomous Agents and Multi-Agent Systems*, 20(1), 80–84.
- Nakaya, K., & Murota, M. (2013). Development and evaluation of an interactive English conversation learning system with a mobile device using topics based on the life of the learner. *Research and Practice in Technology Enhanced Learning*, 8(1), 65–89.
- Nass, C. I., Steuer, J. S., & Tauber, E. (1994). Computers are social actors. *Proceeding of the CHI '94 Proceedings*, 72–78.
- Ockert, D. (2012). Do Japanese JHS students have ideal L2 selves? Evidence from research in progress on the influence of multimedia use on affective variables. In A. Stewart & N. Sonda (Eds.), *JALT2011 conference proceedings*. Tokyo: JALT.
- Ohmaye, E. (1998). Simulation-based language learning: An architecture and a multi-media authoring tool. In R. C. Schank (Ed.), *Inside multi-media case based instruction* (pp. 1–102). Mahwah: L. Erlbaum Associates.
- Öz, H., Demirezen, M., & Pourfeiz, J. (2015). Willingness to communicate of EFL learners in Turkish context. *Learning and Individual Differences*, 37, 269–275.
- Park, S., & Catrambone, R. (2007). Social facilitation effects of virtual humans. *Human Factors*, 49(6), 1054–1060.
- Peng, J. (2007). Willingness to communicate in an L2 and integrative motivation among college students in an intensive English language program in China. *University of Sydney Papers in TESOL*, 2(1), 33–59.
- Poulisse, N. (1987). Problems and solutions in the classification of compensatory strategies. *Second Language Research*, 3(2), 141–153.
- Raux, A., & Eskenazi, M. (2004). Using task-oriented spoken dialogue systems for language learning: potential, practical applications and challenges. In *InSTIL/ICALL Symposium 2004*.
- Reinders, H., & Wattana, S. (2014). Can I say something? The effects of digital game play on willingness to communicate. *Language Learning & Technology*, 18(2), 101–123.
- Rickel, J., & Johnson, L. (1998). Animated pedagogical agents for team training. In *Proceedings of the ITS Workshop on Pedagogical Agents* (Vol. 75, p. 77). San Antonio, Texas.
- Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. *Analyzing Discourse: Text and Talk*, 71, 93.

- Seliger, H. W. (1977). Does practice make perfect? : A study of interaction patterns and L2 competence. *Language Learning*, 27, 263–278.
- Sick, J., & Nagasaka, J. P. (2000). A test of your willingness to communicate in English (Japanese version): Unpublished questionnaire.
- Smith, C., Crook, N., Dobnik, S., Charlton, D., Boye, J., Pulman, S., & Gambäck, B. (2011). Interaction strategies for an affective conversational agent. *Presence: Teleoperators and Virtual Environments*, 20(5), 395–411.
- Suleman, R. M., Mizoguchi, R., & Ikeda, M. (2016). A new perspective of negotiation-based dialog to enhance metacognitive skills in the context of open learner models. *International Journal of Artificial Intelligence in Education*, 26(4), 1069–1115.
- Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64–81). Cambridge: Cambridge University Press.
- Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16(3), 371–391.
- Thornbury, S. (2005). *How to Teach Speaking*. Essex: Pearson Education.
- Van Mulken, S., André, E., & Müller, J. (1998). The persona effect: How substantial is it? In *People and computers XIII* (pp. 53–66). London: Springer.
- VanPatten, B., & Benati, A. G. (2015). Key terms in second language acquisition. Bloomsbury Publishing.
- White, S. (1989). Backchannels across cultures: A study of Americans and Japanese. *Language in Society*, 18(1), 59–76.
- Wolf, J. P. (2008). The effects of backchannels on fluency in L2 oral task production. *System*, 36(2), 279–294.
- Yashima, T. (2002). Willingness to communicate in a second language: The Japanese EFL context. *The Modern Language Journal*, 86, 54–66.
- Yashima, T., Zenuk-Nishide, L., & Shimizu, K. (2004). The influence of attitudes and affect on willingness to communicate and second language communication. *Language Learning*, 54(1), 119–152.
- Yngve, V. (1970). On getting a word in edgewise. In: Papers from the sixth regional meeting of the Chicago Linguistic Society, 567–578.