PAPER REF: 4062

A FRAMEWORK FOR INTERACTING WITH ARTIFICIAL SYSTEMS

Peer M.Sathikh(*), S.G. Lee[2]

[1]School of Art, Design and Media (ADM), Nanyang Technological University, Singapore
[2]School of Mechanical & Aerospace Engineering (MAE), Nanyang Technological University,
(*)*Email:* peersathikh@ntu.edu.sg

ABSTRACT

Everyday, humans interact and execute transactions with artificial systems such as the ubiquitous ATM machine, general ticketing machines (GTMs) at the subway station, customer service kiosks at the airport, to name a few. With mobile communication technology making leaps and bounds, such transactions should be designed from the human-centred point of view for the transaction to be effective. However, such human-to-artificial communication, in most cases, is designed by technical experts rather than user interface designers.

This paper outlines how two foremost communication theories - *Speech Act Theory* and the *Theory of Communicative Action*- may be applied to design a framework for human-to-artificial interactions. This framework is based on five principles: establishing the user's intention and the context in an artificial transaction; the use of directive language; minimizing indirect speech; simplifying illocutionary forces and assuring trust. This framework was tested on four artificial systems; a bank ATM, a GTM, a visitor registration kiosk in a hospital and a customer feedback kiosk in an airport, The analysis of the results led to improvements to the interaction architecture based on speech act theory and theory of communicative action and to the screen graphics design to offer user friendly interaction in each case. This framework could be the starting point for developing a human-centric protocol for interacting with more complex artificial systems, including robots, in the future.

Keywords: human-to-artificial interaction, speech act theory, theory of communicative action.

INTRODUCTION

With the rapid absorption of technology and human-machine interactions becoming ubiquitous, more and more essential services will inevitably be dispensed by artificial systems. However, much of the interaction aspects of these artificial systems are designed by technical experts who are not familiar with human cognitive psychology, leaving users to adapt to these machines, often resulting in miscommunication and undue anxiety, especially for the less tech-savvy. This opens up an opportunity for the study and research into possible improvements to the existing paradigm of interaction between humans and artificial systems, especially when even day-to-day devices such as smartphones, tablets and other 'ubiquitous' gadgets are becoming more and more dependent on a two way interaction between humans and the device or gadget. The intimacy of this interface between the human person and the 'artificial system ' has made it '*impossible to distinguish technology from the social and cultural business of being human* (Tofts, *et al*, 2002)'. Today, this intimacy is more or less controlled by the programming in the gadgets rather than humans (users). Hence, there is a

need for a framework that could lead to natural interaction between humans and machines. This paper explains a brief overview of the landscape of human to artificial interaction in order to establish a need for the understanding of the fundamental concepts behind human-to-human *(*H2H*)* communication before proposing a framework for human to artificial (H2A) interaction. A brief description of the application of the H2A interaction and the findings will then be presented, ending with possible research directions in the future.

EXISTING LANDSCAPE

The interaction landscape today is marked by three different thrusts taking place around the world:

1. The first is the research and development of H2A interaction systems over the last forty years starting from the command and control interface through prompted and/or interactive responses using graphics and voice, graphics user interface of the PCs, PDAs, mobile phones, etc., to natural language interfaces right up to the Universal Speech Interface (USI) developed at Carnegie-Mellon University. Examples of these initiatives may be found in the works of Rosenfield *et al* (2000), Tomko *et al* (2004).

2. The second thrust is a whole body of work inter-personal communications and interactions. This field of work has identified and documented theories on practices, nuances and attributes of H2H communications and relationships. Social and cultural aspects of inter-personal communication are also being studied e.g. Hancher *et al* (1979), Winograd *et al* (1986) and Auramaki *et al* (1988)

3. The third thrust is from a group of thinkers, sociologist and others who investigate into the inter-twining of technology into the everyday life of humans. They are looking at attributes that differentiate a human from an ‘artificial intelligence’ and how these should relate to future products and systems, when it comes to ‘human-to-artificial system’ interaction. Their thoughts can be read in the works of Goldberg (2000), Marsh (1988), Norman (1993), and Searle (1969).

All these three streams of thought and research, while not progressing independently of each other, do not seem to come together to identify a common, balanced approach to ‘H2A’ interaction. In the heart of any H2A interaction is the ‘user interface’, be it as touch screen, mouse, gesture or any other means by which humans could interact and communicate with the artificial system. Commenting on this, in an article titled ‘From Computing Machinery to Interaction Design’, Winograd (1997) says that “the design role is the construction of the "interspace" in which people live, rather than an "interface" with which they interact. The interaction designer needs to take a broader view that includes understanding how people and societies adapt to new technologies”. Winograd (30) in the same article goes on to say, “Interaction design in the coming fifty years will have an ideal to follow that combines the concerns and benefits of its many intellectual predecessors. Like the engineering disciplines, it needs to be practical and rigorous. Like the design disciplines, it needs to place human concerns and needs at the centre of guiding design; and like the social disciplines, it needs to take a broad view of social possibilities and responsibilities”.

DEVELOPMENTS IN H2A INTERACTION

While the touch screen technology used in mobile phones today represent the latest in (graphic) user interface, there are other significant research work being carried out in top universities in the world that are important to H2A interaction.

Universal Speech Interface

Rosenfeld and his team (2000) at Carnegie Mellon University have been developing a Universal Speech Interface (USI) based on a speech recognition algorithm since 2000. Also called Speech Graffiti, the USI is, according to Rosenfeld and his team "an attempt to create a standardized, speech based interface for interacting with simple machines and information servers. Such standardization offers several benefits, including domain portability, lower speech recognition error rates and increased system transparency for users". Though proven to be successful, Speech Graffiti's effectiveness is dependent on 'shaping spoken input' since research showed that users often have trouble speaking within the bounds of its subset language grammar. Shaping spoken input implies a certain level of training for the user to speak within the bounds of the systems grammar.

Organic Interface

Zue (2007) at MIT Computer Science and Artificial Intelligence laboratory argues that "we are not likely to succeed until we can build interfaces that behave more like organisms that can learn, grow, reconfigure, and repair themselves, much like humans". While discussing the input modalities in such an organic interface, Zue goes on to say "in daily interactions, we often rely on pointing, gesturing, and writing to augment speech. There are certainly occasions when speech would not be appropriate, as when we attempt to take notes during a meeting. To provide a full range of interactions and add redundancy, modalities such as pen and gesture should be included to augment and complement speech". On the output side, Zue argues "a multimodal interface must be able to generate natural speech and integrate it in real-time with facial animation, in the context of a larger conversation ".

USI and Organic Interface have been discussed here since the authors feel that both of them have shown direction that seem logical in moving towards H2A Interaction for the future. However both seem have shortcomings in relations to a more natural interaction with the artificial. Both USI and Organic Interface talk about developing algorithms that define their 'grammar' that the user can easily adapt to rather develop a framework or protocol that defines how the artificial should be interacting with humans. In doing so, the user tries to gain a level of 'trust' in the interface the artificial system understands the trained syntax of the user. This trust relies on the hope that any accent and tonal variation, while interacting with the system, does not cause major problems or errors. This approach of users training to a new syntax and relying on the artificial to catch accent and tones brings level of instability in the interaction system.

HUMAN-TO-HUMAN COMMUNICATION

Long before machines/products came along, human kind had been communicating with one another. From basic speech, complex language systems evolved, each with its own grammar and rules. Throughout history, despite the invention of the telegraph, telephone, fax, email and other modern communications of today, people have relied on face-to-face communication when it came to exchanging highly complex information. Face-to-face communication is mostly based on speech, which however requires visual and verbal (aural) cues for the interaction (conversation) to be effective. Where face-to-face interaction was not possible, written form of the language was introduced to communicate with one another and with larger audience. In other words, the written form of language started emerging as languages matured bringing in another facet to human communication in situations where direct communication was not possible. Through verbal and written language, people exchange information, express intention and commit themselves to some action. Over the centuries, trust in communication and interaction was built around a framework for natural communication. In order to bring a level of trust into this form of human-to- human interaction, there was a need for underlying guiding principles for speech and the associated act of communication, which form the foundation of a human to human communication.

Choudhury, *et al* (2006) point out that "as social animals, people's interactions with each other underlie many aspects of their lives: how they learn, how they work, how they play and how they affect the broader community. Understanding people's interactions and their social networks will play an important role in designing technology and applications that are 'socially-aware'". While a completely social artificial system may be many years away, the first step towards a socially interactive artificial system could be based on the communication medium of humans, namely speech.

Searle's (1969) *Theory of Speech Acts* and Habermas' (1981) *Theory of Communicative Actions*, which is an extension of the *Theory of Speech Acts*, are arguably the foundations of this field of study into human interaction. Language Action Perspective (LAP) outlined by *Theory of Speech Acts* explains how language coordinates communications among people, assuming a common ontology (and trust) exists among communicating parties,. The trust could be said to exist in a speech act bring about an appropriate communicative action or actions that is recognized by both the parties as a result of their interaction. Huh? H-E-L-P!!

Social semiotics examines the influence of signs in the context of society and culture in H2H communication and comprises three aspects – *semantics*, *syntactics* and *pragmatics*. Both *semantics* and *syntactics* are linguistic-biased, being more concerned with the language structure, whereas *pragmatics* takes into account the *context* of communication, for instance, the status of the speaker and hearer, the inferred intent of the speaker, and other factors. An awareness of social semiotics underpins most H2H communication in *Speech Act Theory* and the *Theory of Communicative Action.*

In the course of research towards establishing a framework for H2H interaction, both the *Speech Act Theory* and the *Theory of Communicative Action* were studied in detail and a framework was established for H2A interaction. Four existing common H2A systems, namely a bank ATM, a General Ticketing Machine (GTM), a registration kiosk in a hospital and a customer feedback kiosk were studied based on the prevailing principles of human-to-machine interaction. A new framework for H2A interaction was then proposed, based on

Speech Act Theory and the Theory of Communicative Action, to establish an interaction flow and information architecture.

SPEECH ACT THEORY

Oxford philosopher and linguistic theorist J.LAustin argued that when people say something, they are not merely saying 'something' but rather intending for that something to happen [15]. This desire is termed a *speech-act* by Austin which his student Searle (1969) went on to develop further. *Speech Act* has three components: a *locutionary* act; an *illocutionary act*; and a *perlocutionary* act as shown in Figure 1.

A *locutionary act* is a phonetic, syntactic and semantic utterance. An illocutionary act, on the other hand, expresses the speaker's intent or 'attitude' towards some propositional content. This intent of the speaker, be it to inform, make a request, effect a change or to express a personal feeling, is conveyed in an *illocutionary force*. An illocutionary force, coupled with some propositional content, constitutes an illocutionary *act*. Therefore, illocutionary forces underpin the overall success of a speech act. A *perlocutionary act* conveys the speaker's intention to the addressee through the illocutionary act. A perlocutionary act has succeeded if the addressee executes the illocutionary act. Therefore, in analyzing H2M interaction, all illocutionary acts must be expressed explicitly and unambiguously for the H2M communication to succeed.

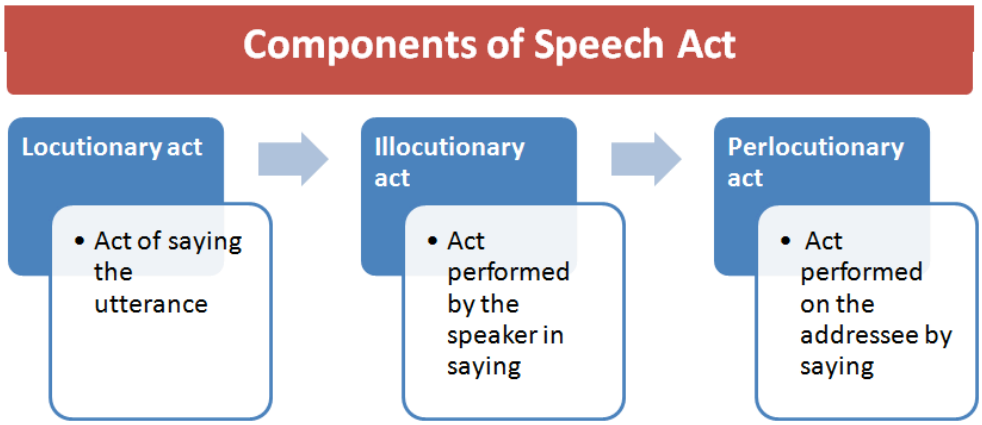

Fig.1 Components of Speech Act

Searle further identified five fundamental intentions of the speaker or illocutionary 'forces' as shown in Table 1. The illocutionary point of an utterance not only relies on the semantic and syntactic meaning of the utterance but also the shared contextual background of both the speaker and the addressee. If an utterance has different illocutionary forces embedded in it, it is called an *indirect speech act*. Indirect speech acts are commonly used to make a request or to reject a proposal in H2H communication. To decipher the primary illocutionary force of an indirect speech act, one has to infer from background information. This is a hindrance to communication and especially so for H2M communication since pre-programmed interactive devices often lack the ability to make inferences.

Table 1 Fundamental Intentions of the speaker

	Categorization	Illocutionary point	Example
1	**Assertives**	Commit speaker to the truth of the expressed proposition	**Informing** "Your transaction is complete"
2	**Directives**	Attempt by the speaker to get the hearer to do something	**Requesting** "Please take you card"
3	**Commissives**	Commit the speaker to some future course of action	**Promising** "We pay 3% interest"
4	**Expressives**	Express the psychological state about a state of affairs specified in the propositional content.	**Apologizing** "We apologize for the inconvenience caused"
5	**Declaratives**	To effect a change which brings about the correspondence between the propositional content and reality	**Invalidating** "Your card has been cancelled after 3 Invalid PIN entries"

THEORY OF COMMUNICATIVE ACTION (TCA)

Habermas (1981) postulated in his *Theory of Communicative Action*, that *Speech Act Theory* does not deal with the orientation of rational participants towards mutual agreement. *Speech Act Theory* deems a speech act successful if the desired course of action implicit in the perlocutionary act is achieved. Habermas, however, countered by suggesting that the success of a speech act not only depends on whether the hearer understands the speaker, but that she accepts that the speaker has the authority and is sincere, that the proposition is feasible and that the speech act is valid in the context . Habermas classified these into four 'claims: power, sincerity, truth and justice.

A PROPOSED FRAMEWORK FOR H2M COMMUNICATION

Many researchers have applied the two theories essentially to information systems rather than to H2A interaction. The authors' propose a framework for H2M interaction based on the Speech Act Theory to ensure that speech acts initiated by an artificial system are comprehensible, after which the *Theory of Communicative Action* verifies all the validity claims. The framework is characterized by five actions as shown in Table 2.

Table 2 Fundamental Framework

	Key attribute	Elaboration
1	Who are the users?	Familiar users understand the machine-initiated speech acts due to prior experience. Not so for new users who need more propositional content.
2	What are their intentions?	The machine-initiated speech acts and their sequence must fulfill these intentions. The illocutionary forces and propositional content of the speech acts then can be formalized into machine functions.
3	Is the propositional content clear?	Sometimes, the user cannot interpret or mis-interprets the speech act. In pre-programmed H2H interaction, all propositional content must be unambiguous. Since new users are most prone to this, a 'speech bubble' can possibly help.
4	Is the illocutionary force clear?	The primary illocutionary force is the intended illocutionary force and is often a directive while the secondary illocutionary force is often an assertive. Employ unambiguous *directives* which motivate the user to execute the intended perlocutionary act, backed up by *assertives* which verify the validity claims. Supply adequate cues to guide the user. Avoid the use of *indirect speech acts* altogether.
5	Are the claims valid?	A speech act is successful only when the user is convinced of the *validity* of the request. For interactive machine systems, claims to *truth* and *sincerity* of the speech acts are unlikely to be challenged. However, the user may not be convinced of the claim to *justice* when the machine requests private information. In such an event, the machine has to win over the user.

BASIC VALIDATION OF THE AUTHORS' PROPOSED H2M INTERACTION FRAMEWORK

The authors decided to test the validity with simple artificial systems which were different types of interactive self-service machines. The selected artificial systems, namely a bank ATM, a subway General Ticketing Machine (GTM), a registration kiosk for visitors to a hospital as well as a passenger feedback kiosk of an international airport as shown in Table 3.

Table 3 Artificial systems for validation

Service	Description	Examples
Simple	Financial transaction w/o need to authenticate the user	GTM, vending machines
Complex	Financial transaction with need to authenticate the user	Bank ATM
Registration	Collects personal data and saves in a central database	Hospital registration kiosk
Data gathering	Collects general information from respondents	Airport feedback kiosk

The authors' proposed framework was validated against the four interactive self-service machines, revealing the shortcomings and proposed remedies as detailed in Table 4.

Table 4 Specific Improvements to the Framework

Service	Issues	Proposed rectification
ATM	• The option "service menu" shows cash denominations "$50", "100", "$1000" but the *propositional content* is inadequate as the verb "withdraw" is missing. The option to personalize one's transactional preferences through "My ATM" also lacks propositional content. • "Minimum withdrawal is $20" is an *indirect speech act*. Its intended objective is a directive (to get the user to withdraw more than $20) while the secondary illocutionary force is an assertive (i.e. that the minimum amount is $20). • The user is informed that his intention cannot be realized after 3 speech acts (insert card, enter PIN, select account).	More detailed propositional content is needed but should be balanced between the needs of familiar and novice users. To eliminate indirect speech act, "Minimum withdrawal is $20" should consist of 2 separate speech acts; a directive "Please withdraw more than $20" and a supportive assertive "Minimum withdrawal is $20". Inform user early if his intentions cannot be met, by proper sequencing of the machine-initiated speech acts as early as possible in the interaction.
GTM	• The main function of the GTM is to dispense subway tickets. As unfamiliar users such as tourists are most likely to use it, the directive speech acts must be strong. E.g. "Select destination" does not direct	Stronger directives are needed.

	the user to where he can select the destination; "Please collect your standard ticket" does not indicate where. • The validity of the speech acts is not likely to be challenged.	
Registration kiosk	The purpose of the kiosk was for visitors to the hospital wards to self-register so they can be contacted in the event of an epidemic and to regulate the number of visitors. • No assertive information on (i) one-off registration each day; (ii) scanning ID, and (iii) getting the name, bed and ward numbers of the patient.	The kiosk should cater to both new and repeat visitors through more assertive speech acts. Request for the patient's details first before asking for the visitor's. Make the propositional and the illocutionary forces of the speech acts comprehensible. The validity claims of the visitor have to be addressed by explaining why the information is requested.
Feedback kiosk	The objective of the feedback kiosk is to survey airport users feedback on the airport's facilities and services. However, the response rate had been dismally low. Some reasons include • The kiosk design is dull and un-inspiring and passers-by do not know that it is a feedback kiosk. • The user is asked to rate seven (7) aspects of airport service. This can deter passers-by who are in a hurry. • There are too many data-entry fields for the user to fill up. • After filling in his personal details, the user is asked to return to the feedback screen.	• A stronger directive is needed to draw attention to it. • Provide a "Quick feedback" option. • Indicate which data fields are mandatory and which are optional. • The sequence of the speech acts should foster more logical user interaction.

OUTCOME OF THE VALIDATION EXERCISE

Table 5 summarizes how each artificial system stacks up against the key attributes of the H2A interaction framework.

The GTM complies best to the principles of the authors' H2M interaction framework. It understood the users' intention and systematically sequenced the speech acts to help

unfamiliar users. Its drawback is the absence of strong directives, which is a major setback as the machine caters to a wide range of users, from the familiar to new users such as tourists.

The bank ATM arguably does not comply with the H2M interaction framework. However, it makes use of strong directives and fulfills all validity claims.

Table 5 Comparing the system with H2A framework

Interactive machine	Key attributes of the framework for H2A interaction					
	Caters to main user?	Understands user intentions?	Clear propositional contents?	Clear illocutionary forces?		Validity claim?
				Avoids indirect speech act?	Strong directives?	
ATM					☑	☑
GTM	☑	☑	☑	☑		☑
Hospital registration kiosk			☑	☑		☑
Airport feedback kiosk	☑		☑	☑		

CONCLUSION

. A framework for human to artificial systems (H2A) interaction based on the Speech Act Theory and the *Theory of Communicative Action* is proposed. The framework is based on the fundamental semantic and syntactic intentions of the human being in a given contextual background. The authors' proposed H2A interaction framework was validated against four interactive systems. The General Ticketing Machine best complies overall to the proposed framework although it does not make use of strong directives. The bank ATM however has strong directives but falls short of the proposed framework.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contribution of Ms S H Lee and Ms A G Dhamarwan.

REFERENCES

Auramaki, E., Lehtinen, E., & Lyytinen, K. (1988). *A speech-act-based office modeling approach*. ACM Trans on Info Systems. Vol 6.

Auramäki E, & Lyytinen, K. (1996). *On the success of speech acts and negotiating commitments.* In Proc of the 1st Int workshop on the Language Action Perspective´96, Springer Verlag 32.

Choudhury, T., Philipose, M., Wyatt, D., Lester, J. (2006). *Towards Activity Databases: Using Sensors and Statistical Models to Summarize People's Lives*. IEEE Data Engineering Bulletin. March 2006.

Habermas, J. (1981). *The Theory of Communicative Action.* Translated by Thomas McCarthy. Boston: Beacon Press

Janson, M. A., Woo, C. C., & Smith, L. D. (1993). *Information systems development and communicative action theory*. Information & Management, *25*(2), 59-72.

Kumar, K., & Becerra-Fernandez, I. (2007). *Interaction technology: Speech act based information technology support for building collaborative relationships and trust*. Decis. Support Syst., *43*(2), 584-606.

Rosenfield, R., Olsen, D., Rudnicky, A., (2000). *Universal Human-Machine Speech Interface : A White Paper (CMU-CS-00-14)*. Carnegie-Mellon University.

Searle, J R. (1969). *Speech acts: an essay in the philosophy of language*. London: Cambridge University Press.

Searle, J.R (1979). *Expression and Meaning: Studies in the Theory of Speech Acts*. London: Cambridge University Press.

Tofts, D., Jonson, A., Cavallaro, A. (Eds.) (2002). Prefiguring Cyberculture: an intellectual history. MIT Press.

Tomko, S. & Rosenfeld, R. (2004). Speech Graffiti vs. Natural Language: Assessing the User Experience. Proc. HLT/NAACL, Boston, MA.

Winograd, T., Flores, F. (1986). *Understanding Computer and Cognition : A New Foundation for Design*. Intellect Books.

Winograd, T. (1997). *From Computing Machinery to Interaction Design.* in Denning, P., Metcalfe R.,(Eds.), Beyond Calculation: The Next Fifty Years of Computing (pp. 149-162), Springer-Verlag.

Zue, V. (2007). *On Organic Interfaces*. INTERSPEECH-2007, 1-8.