ICED13
**INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED13**
19-22 AUGUST 2013, SUNGKYUNKWAN UNIVERSITY, SEOUL, KOREA

# A FRAMEWORK FOR EFFECTIVE HUMAN-TO-MACHINE COMMUNICATION FOR ARTIFICIAL INTERACTIVE SYSTEMS

**Siang Guan LEE, Peer SATHIKH**
Nanyang Technological University, Singapore

## ABSTRACT

Most artificial interactive systems are designed by technical experts and not from the point of view of human-to-machine (H2M) communication.
This paper discusses how two foremost human-to-human (H2H) communication theories – Speech Act Theory and the Theory of Communicative Action, are employed to propose a framework for the design of interactive artificial systems.
The proposed framework embodies five attributes: the user's intentions and the context of the transaction; employing strong directive language while minimizing indirect speech, and assuring trust.
The framework was tested on four interactive machines: a bank ATM, a subway ticketing machine, a registration kiosk & a customer feedback kiosk. The GTM best complied with the attributes of the framework: the ATM the least.

*Keywords: user integration, speech act theory, theory of communicative action*

Contact:
Dr. Siang Guan Lee
Nanyang Technological University
School of Mechanical & Aerospace Engineering
Singapore
639798
Singapore
msglee@ntu.edu.sg

# 1 INTRODUCTION

Human-machine interactions are becoming ubiquitous in today's fast-paced life. As self-service equipment are efficient and cost effective, more services will inevitably be dispensed by machines. NCR Corp., the maker of automated teller machines, achieved a record sale of its self-service check-out machines in 2010 and expects growth of 10 percent or more in 2011 (Credeur, 2011). Large retailers including Wal-Mart, IKEA and Tesco are slowly transforming their manned check-out lanes into self-service ones (Miletic, 2008). Meanwhile in the travel sector, IATA, which represents 93 per cent of the world's airlines, has proffered what it calls a *Fast Travel* manifesto – a plan to transform airports into complete self-service operations (Archer, 2010). Through *Fast Travel*, passengers scan their own passports, check themselves and their bags in through self-service kiosks.

However, as these artificial systems are designed by technical experts who are not familiar with human cognitive psychology, the human-to-machine (*H2M*) interaction is sub-optimal. The user needs to adapt to these machines, often resulting in miscommunication and undue anxiety. Hence, there is a need to improve the interaction between humans and machines. This paper presents an overview of the landscape of human to machine interaction and the fundamental concepts behind human-to-human *(*H2H*)* communication before a framework is proposed for human to machine (H2M) interaction. Four common interactive machines were analyzed by the framework, and the results discussed.

# 2 THE EVOLUTION OF H2M INTERACTION

Over the last 40 years or so, developments in H2M interaction are marked by two different thrusts:

1. The first is H2M interaction interfaces 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. Other initiatives may be found in the works of Rosenfield *et al* (2000) and Tomko *et al* (2004).
2. The second thrust is in the social and cultural aspects of inter-personal communications and interactions e.g. Hancher *et al* (1978), Winograd *et al* (1986) and Auramaki *et al* (1988).

However, a common, balanced approach to H2M interaction is still very much work-in-progress. Although the user interface is crucial to effective H2M interaction, Winograd (1997) pointed out that "the design of the "interspace" in which people live, rather than an "interface", is more important, as people and societies adapt to new technologies".

# 3 DEVELOPMENTS IN H2M INTERACTION

While the touch screen technology used in mobile phones today represent the latest in (graphic) user interface, other significant research into H2M interaction carried out worldwide include the following.

Rosenfield *et al* (2000) of Carnegie Mellon University developed a Universal Speech Interface (USI) based on a speech recognition algorithm. *Speech Graffiti* is "an attempt to create a standardized, speech-based interface for interacting with simple machines and information servers". Though successful, *Speech Graffiti's* effectiveness is dependent on "shaping spoken input so users learn to speak within the bounds of its subset language grammar."

Zue (2007) at MIT's Computer Science and Artificial Intelligence Laboratory argues that "we should build interfaces that behave more like organisms that can learn, grow, reconfigure, and repair themselves, much like humans". Zue argues that "a multi-modal interface is needed that can generate natural speech and integrate it in real-time with facial animation, in the context of a larger conversation ". Which was what University of Cambridge researchers have developed, namely a virtual "talking head" that is capable of expressing six basic human emotions– happiness, sadness, fear, anger, tenderness and neutrality, as well as changeable pitch, speed and depth settings (Cambridge Talking Head, 2013). The user first enters a line of text and then moves a range of sliders to determine emotion. After the Enter key is pressed, the avatar will read the message in the emotion desired. Unlike Apple's *Siri* and Samsung's *S Voice*, the avatar engages in (fairly) life-like conversations with smartphones, making interaction with machines more life-like.

# 4 HUMAN TO HUMAN COMMUNICATION

Before we delve into H2M interaction, let us examine H2H communication. After human kind had learnt to communicate verbally with one another face-to-face, a written form of the language later emerged. Over the centuries, the foundation of human to human communication evolved, underpinned by principles of speech and communication. 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 human interaction. *Language Action Perspective* (LAP) an offshoot of *Theory of Speech Acts* explains how language coordinates communications among people, assuming a common ontology (and trust) exists among communicating parties.

Another facet of H2H communications deserves mention; *Social semiotics* which has 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 and the inferred intent of the speaker. An awareness of *pragmatics* underpins most H2H communication in *Speech Act* and *Communicative Actio Theories* (Blakemore, 1992). In the course of research towards establishing a framework for H2H interaction, both theories were studied in detail and a framework was established for H2M interaction.

## 4.1 Speech Act Theory

The linguistic theorist Austin argued that when people say something, they are not merely saying 'something' but rather intending for that something to happen (Austin, 1975). This desire is termed a *speech act* by Austin. A *Speech Act* has three components: a *locutionary* act; an *illocutionary act*; and a *perlocutionary* act (Austin, 1975), as shown in Figure 1.

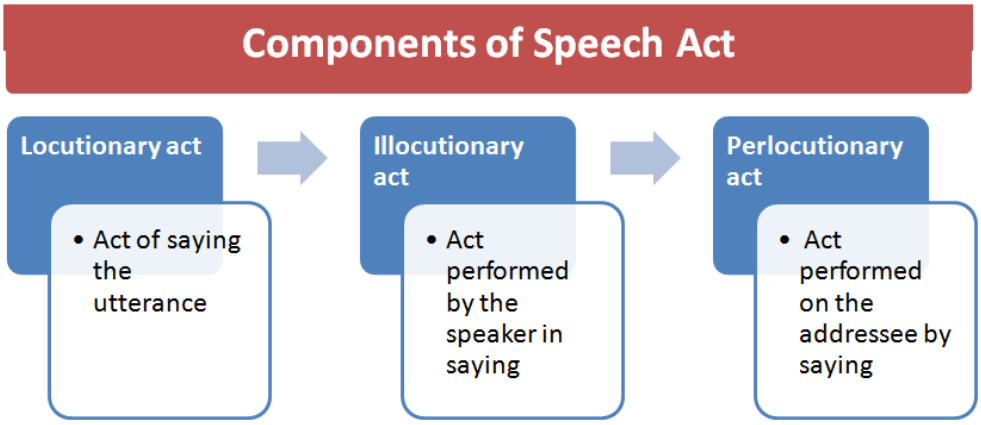


*Figure 1: The three components of Speech Act*

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 (Verschueren & Östman, 2009). 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, constitute 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 (Habermas, 1981). A perlocutionary act succeeds if the addressee executes the illocutionary act. Therefore, in H2M interaction, all illocutionary acts must be expressed explicitly and unambiguously for the H2M communication to succeed.

Searle (1975) identified five fundamental intentions of the speaker or *illocutionary points* (see Figure 2). 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, especially so for H2M communication since pre-programmed interactive devices often cannot make inferences. Therefore in H2M communication indirect speech acts are to be avoided.

| | 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" |

*Figure 2: The five fundamental intentions of a speaker*

### 4.2 Theory of Communicative Action (TCA)

Habermas postulated his *Theory of Communicative Action* (TCA) to address the orientation of rational participants towards mutual agreement where *Speech Act Theory* does not (Janson & Woo, 1995, Habermas, 1981, Dietz & Widdershoven, 1991). *Speech Act Theory* deems a speech act successful if the desired course of action implicit in the perlocutionary act is achieved. Habermas suggested that the success of a speech act not only depends on whether the hearer understands the speaker, but that she accepts that the speaker is sincere and has the authority, that the speech act is valid in the context and that the proposition is feasible. Habermas classified these into four claims: power, sincerity, truth and justice. As Habermas's TCA is hearer-oriented, it has the potential to deepen understanding of H2M communication.

## 5 A PROPOSED FRAMEWORK FOR H2M COMMUNICATION

Many researchers have applied the two theories essentially to information systems. Some (Ljungberg & Holm, 1996; Suchman 1994) claimed that Habermas´ theory is superior to Searle's theory of H2H communication but are skeptical about Habermas' validity claims. Other researchers amalgamated both theories: Auramäki & Lyytinen (1996) and Schoop (1999) articulated H2H communication in which the speaker strives to reach an understanding with the hearer. The authors have distilled what they consider to be most relevant in both theories and propose a framework for H2M communications. Table 1 summarises the authors' framework which is founded on Illocutionary Forces, Propositional Content, Validity Claims and other General Guidelines.

*Table 1: The authors' H2M communications framework.*

| **Guidelines to Good User Interface Design for Interactive Machines** | | |
|---|---|---|
| **A. Guidelines for Illocutionary Forces** | | |
| **PROCESS** | **HOW?** | **WHY?** |
| 1. Beginning | Start with an expressive | To greet/welcome the users |
| 2. End of each screen | Use directive; except for showing processing and ending | To always tell users what to do next |
| 3. Show system is processing | Use assertive | To tell the users the state of the machine |
| 4. Ending | Use expressive | To give respect to users |
| **B. Guidelines for Propositional Contents** | | |
| 1. Cater both new and | Use minimum number of words; make | To make the speech more |

| | | |
|---|---|---|
| familiar users | use of speech bubble, question mark icon, or pop-ups for more explanation | understandable |
| 2. Presentation of choices | Put all the choices in one page clearly and direct the users to choose | To make the choices clearer to the users |
| 3. Use of directive verb | Use unambiguous directive | To make the speech more understandable |
| 4. Choice of words | Do not use jargon or complicated sentences; common vocabularies are preferred | To make the speech more understandable |
| **C. Guidelines for Validity Claims** | | |
| 1. Identify user's intention | From part B.2, the end of a chain of choices will determine the user's main intention | To ensure the purpose of the user is achieved in the end |
| 2. Identify speech acts that could raise validity claim questions | Usually those that are not in line with user's main intention identified in C.1 | To be able to answer and reduce validity claim questions |
| 3. Identify the validity claim questions | Ask claim to rightness (context) and truthfulness (reason) to the speech acts | To be able to answer and reduce validity claim questions |
| 4. Answer the validity claim questions | The validity of questionable speech acts can be substantiated and enhanced by other types of illocutionary force | To establish mutual agreement with the user. |
| **D. Other General Guidelines** | | |
| 1. Avoiding indirect speech act | Break the sentences to show all the forces intended | To minimize misunderstanding and confusion |
| 2. Beginning | Tell the users what things they are required to prepare | To increase efficiency |
| 3. Throughout the process | Use sequence bar; create as short a sequence as possible in realising user's main intention | To show the users the general view of whole process and increase efficiency |
| 4. Showing how | When the directives require physical actions, show users how. Best mean: use video | To make the speech more understandable |

At the same time, many different types of interactive self-service machines have been developed over the years. They can be broadly categorized into four types as shown in Table 2.

*Table 2: Four general types of interactive self-service machines.*

| Service | Description | Examples |
|---|---|---|
| **Simple** | Financial transaction w/o need to authenticate the user | General Ticketing Machine (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 chose to analyze these machines, starting with the ATM.

### 5.1 Automated Teller Machine (ATM)

Figure 3 depicts the general sequence for a cash withdrawal in a current ATM.
Cash withdrawal was chosen because it is the most frequently used ATM function. As shown in figure 3, cash withdrawal involves several steps. Table 3 summarises the outcome of the analysis of the ATM user interface. The machine gets a score of 1 if it completely satisfies a point, 0.5 if it only partially satisfies, and 0 if it does not satisfy the point at all.

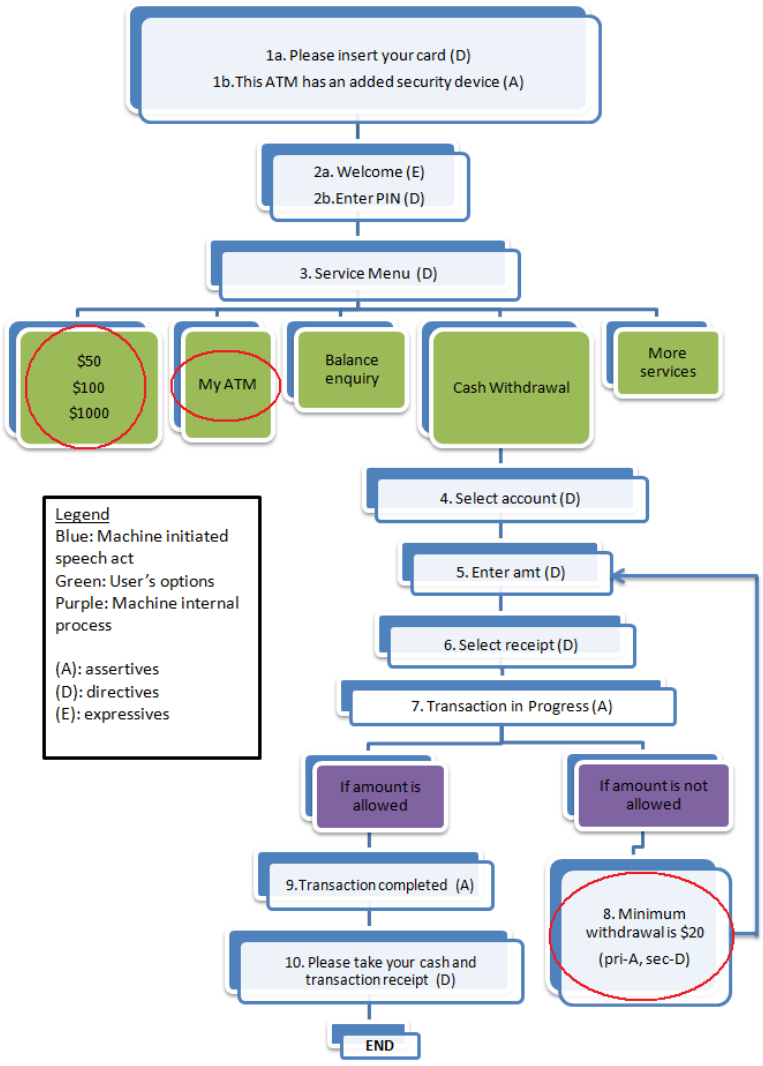


*Figure 3: Process for cash withdrawal*

*Table 3: The outcome of an analysis of the ATM*

| A. Guidelines for Illocutionary Forces | | | |
|---|---|---|---|
| **PROCESS** | **TECHNIQUE** | **CURRENT ATM** | **SCORE** |
| 1. Beginning | Start with an expressive | No | 0 |
| 2. End of each screen | Use directive; except for showing processing and ending | Yes | 1 |
| 3. Show system is processing | Use assertive | Yes | 1 |
| 4. Ending | Use expressive | No | 0 |
| **B. Guidelines for Propositional Contents** | | | |
| 1. Cater both new and familiar users | Use minimum number of words; make use of speech bubble, question mark icon, or pop-ups for more explanation | Doesn't cater to new users, doesn't make use of speech bubble | 0 |
| 2. Presentation of choices | Put all the choices in one page clearly and direct the users to choose | Choices aren't clear | 0 |

| | | | | |
|---|---|---|---|---|
| 3. | Usage of directive verb | Use unambiguous directive | Yes | 1 |
| 4. | Choice of words | Do not use jargon or complicated sentences; common vocabularies are preferred | My ATM is not explained | 0 |
| | C. Guidelines for Validity Claims | | | |
| 1. | Identify user's intention | From part B.2, the end of chain of choices will determine the user's main intention | In this case, we assume cash withdrawal | - |
| 2. | Identify speech acts that could raise validity claim questions | Usually those that are not in line with user's main intention identified in C.1 | No validity claim questions, all directives are in line with user's intention | 1 |
| 3. | Identify the validity claim questions | Ask claim to rightness (context) and truthfulness (reason) to the speech acts | N.A. | 1 |
| 4. | Answer the validity claim questions | The validity of questionable speech acts can be substantiated and enhanced by other type of illocutionary force | N.A. | 1 |
| | D. Other General Guidelines | | | |
| 1. | Avoiding indirect speech act | Break the sentences to show all the forces intended | Not satisfied. "Minimum withdrawal is $20." | 0 |
| 2. | Beginning | Tell the users what things they are required to prepare | Not satisfied. Users are not told to decide from which account they want to withdraw their money. | 0 |
| 3. | Throughout the process | Use sequence bar; create as short a sequence as possible in realizing user's main intention | Sequence is ok, but sequence bar is not utilised. | 0.5 |
| 4. | Showing how | When the directives require physical actions, show users how. Best mean: use video | Not satisfied. Users might not know where to insert the card, collect cash and receipt. | 0 |
| | | | | **Total Score: 6.5** |

## 5.2 Outcome of the analysis

In summary, the current ATM system scores 6.5 out of a maximum of 15 points. The other three interactive machines were likewise analysed and their absolute scores shown in Table 4.

Figure 4 summarises the major strengths and drawbacks of each machine as analysed by the framework. The GTM 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 least complies with the H2M interaction framework. However, it makes use of strong directives and fulfills all validity claims.

*Table 4: The relative scores of the four interactive machines*

| INTERACTIVE MACHINE | SCORE (OUT OF 15) |
|---|---|
| ATM | 6.5 |
| General Ticketing Machine | 10.0 |
| Hospital Registration Kiosk | 9.5 |
| Airport Feedback Kiosk | 8.5 |

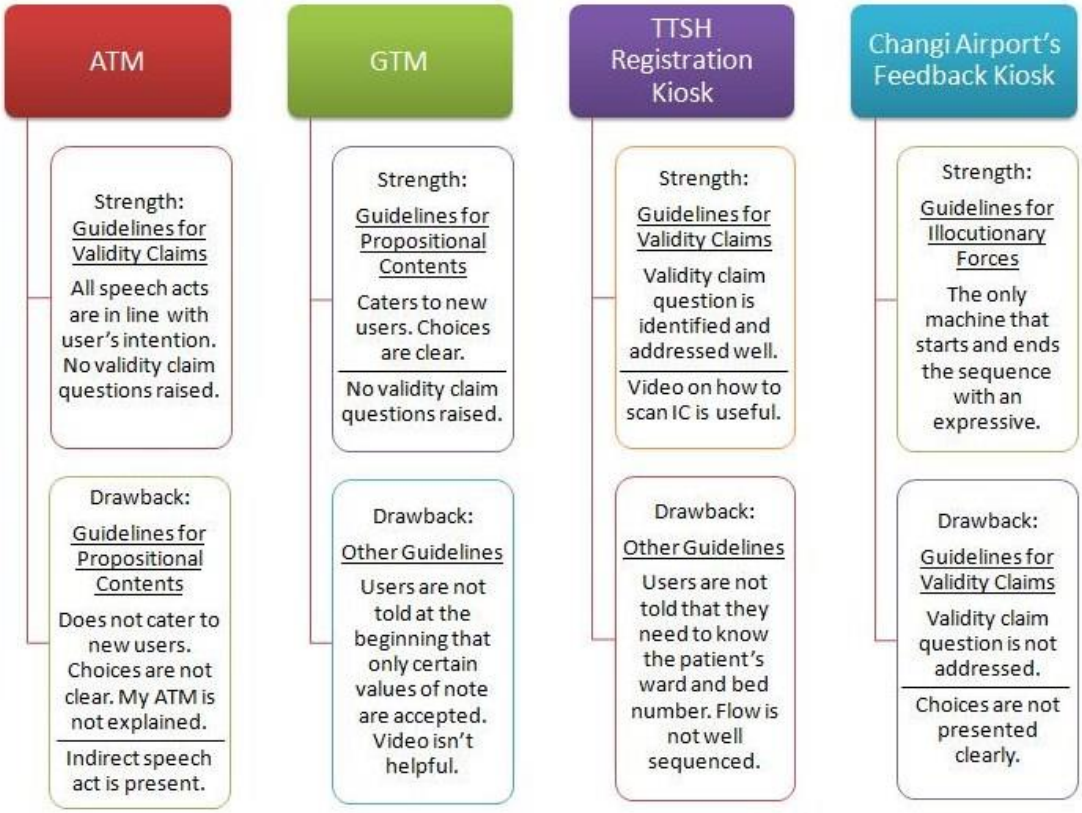


*Figure 4: The key strengths and limitations of the four interactive machines*

## 6 CONCLUDING REMARKS

A H2M interaction framework is proposed based on *Speech Act Theory* and the *Theory of Communicative Action*. SAT postulates that a speaker has an underlying intention that what is proposed will be accomplished; e.g. informing, requesting, making a promise, apologizing, invalidating. TCA embellishes SAT by assuming that the speaker is sincere and has the authority to make the proposition.

The framework was use to analyze four (4) interactive self-service machines: a bank ATM, the subway General Ticketing Machine (GTM), a registration kiosk for visitors to a hospital as well as a passenger feedback kiosk of an international airport.

The GTM complies best with the key attributes of the H2M interaction framework, except for the absence of strong directives. On the other hand, the bank ATM had strong directives and fulfills all validity claims but did not cater to both familiar and new users.


### ACKNOWLEDGEMENTS

The authors are grateful to Ms S H Lee who gathered and analyzed the data.

## REFERENCES


Archer, B. *Longing for the self-serve airport*. The Globe and Mail (Business travel section). Retrieved from http://www.theglobeandmail.com/ (Jul 27, 2010).

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

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

Austin, J L. (1975). *How To Do Things With Words*. NY: Harvard University Press.

Verschueren, J., & Östman Jan-Ola (Eds) (2009) *Key Notions for Pragmatics.* Amsterdam: John Benjamins Publishing Company. pp 232

Bach, K & Harnish, R.M. (1979). *Linguistic Communication and Speech Acts*. Cambridge, MA: M.I.T. Press.

Blakemore, D. (1992) *Understanding Utterances: The Pragmatics of Natural Language*, Oxford: Blackwell.

Cambridge Talking Head. Retrieved from http://www.gizmag.com/zoe-talking-head-avatar/26716/?utm_source=Gizmag+Sub. (26 March 2013)

Cooke, M. (1994). *Language and Reason: A Study of Habermas' Pragmatics*. Cambridge, MA: MIT Press.

Credeur, M. *NCR CEO Forecasts Growth in Self-Service Machines as Customers Cut Costs.* Bloomberg News. Retrieved from http://www.bloomberg.com/ (Jan 25, 2011).

Cruse, D. A. (2004). *Meaning in language: An introduction to semantics and pragmatics*, Chpt , Oxford Textbooks in Linguistics.

Dietz, J. & Widdershoven, G. (1991). *Speech acts or communicative action?* Proceedings of the 2nd European Conference on Computer-Supported Cooperative Work, Amsterdam, The Netherlands.

Goldberg, R.G. & Riek, L. (2000). A practical handbook of speech coders. Boca Raton, Fla: CRC Press.

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

Hancher, M. (1978). "*Describing and Interpreting as Speech Acts*". *Journal of Aesthetics and Art Criticism*. 36: 483-485

Janson, M.A, & Woo, C.C. (1995). *Comparing IS development tools and methods: using speech act theory*. Inf. Manage. **28**(1): 1-12.

Ljungberg J & Holm, P. (1996). *Speech acts on trial.* Scandinavian Journal of information systems. 8(1).

Miletic, D. *A new way to shop — check it out for yourself.* The Age (National section). Retrieved from http://www.theage.com.au/ (Jul 27, 2011)

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

Schoop, M. (1999). *An empirical study of multi-disciplinary communication in healthcare using a language action perspective.* In Proc of the 4th Int Workshop on the Language Action Perspective (LAP99), Copenhagen.

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

Suchman, L. (1994) *Do categories have politics? The Language/Action Perspective reconsidered. Computer Supported Cooperative Work* - an international journal, **2** (3) pp 191-197.

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

Verschueren, J., & Östman Jan-Ola (Eds) (2009). *Key Notions for Pragmatics.* Amsterdam: John Benjamins Publishing Company. pp 232.

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* Springer-Verlag. pp 149-162.

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