Speech Generation by Artificial Intelligent Systems: Issues and Challenges

Mian Khurram Shahzad Azam

Natural Language Processing (NLP) technology has greatly evolved in the last decade. From simple text based processing systems to the emergence of speech comprehension and speech generation systems, natural language processing has shown credible achievements. Artificial Intelligence research has reported notable successes in speech processing technologies in humanoid robots like Kismet (2011) and ASIMO (2011). Yet, there are some basic issues which need to be highlighted in the artificial intelligence research so that meaningful and logical speech comprehension and generation is possible. To develop significant artificial intelligent speech systems for tourism, health, education, industrial and corporate sector, the imperative is to ask if machines can generate speech utterances that reckon with the idea of 'self' in a social and cultural context. For such successes to occur the process of communication has to become central for research in this area and it is critical to explore the human aspect in speech generation and speech comprehension systems.

The paper raises critical questions identified during a comprehensive survey of the existing literature in this area of research and these are: would speech generation systems be able to produce 'creative' utterances? Would these systems still be called creative when they rely on the database of human languages? Serious attention to these questions can give a new direction to the future researchers to look deeply for the development of artificial intelligent

speech comprehension and generation systems that adhere to the socio-cultural aspect of utterances if reliable, logical, meaningful and effective communication act between humans and machines is the goal of speech processing by artificial intelligent systems.

Introduction

In the last two decades, a sizeable amount of research work has been done by researchers like Naoko (1993), Hirschberg et al. (1999), Zue et al (2000), Varchavskaia et al (2001) and Fitzpatrick (2003), to develop an artificial intelligent system that can generate human-like speech as simple speech utterances in a human-machine communication act. The desired outcome of all these research endeavors is to have a logical and meaningful speech act between humans and artificial intelligent systems. To assemble an artificial intelligent communication system which can work better than the human brain in creativity and communication has been a dream of researchers in artificial intelligence. Recent developments in computer science research have achieved success in creating computing systems that are highly efficient in processing data. It appears that gradually the comparison of human capacity with computer performance is losing appeal as faster and more efficient systems are competing with each other for greater speed in performance and processing of data.

The research in artificial intelligence aimed to achieve perfection and capacity beyond human limitations for machines and super-computers. One of the goals of research in artificial intelligence is to focus on creating ability in artificial intelligent systems to use human-like speech with natural ease in communication with human beings. In the last decade, Natural Language Processing (NLP) technology has evolved as a specialized research area which has generated new sets of issues and questions for researchers in artificial intelligence and linguistics. The most immediate issues deal with the

logical link of language database with the processing speed of a system for speech comprehension and speech generation technologies and the maximum ability of the system to correct itself and find quick solutions to resolve ambiguities in speech comprehension and speech utterances. The speech processing systems are required to keep intact the semantic and syntactic value of speech for meaningful and logical speech output.

Transhumanism is the theoretical framework for this research and it is connected with the notable successes achieved in each of the systems/models discussed in the literature review. Transhumanism encourages the technological advancement and research which could take human society beyond the human limitations. Languages, once considered as the sole pride of human beings, are now considered as computable act, for example, Kismet (2011), despite severe opposition from people like Searle (1969) who believed that human languages are not computable and the machines could never know exactly what they are processing as language. By questioning the existing beliefs and giving technology an open chance for finding new avenues, Transhumanism provides the appropriate ground for technological based solutions to human biological, physical and cognitive limitations. It promotes the use of technologies including artificial intelligence and speech comprehension and speech generation technologies for the benefit of human society. The human attribute of language usage for communication is envisaged in artificial intelligent machines and humanoid robots as technological improvement in these machines to use human language for communication with human beings.

Bostrom (2003) defines Transhumanism as:

> The intellectual and cultural movement that affirms the possibility and desirability of improving the human condition through applied reason, especially by developing and making widely available technologies to eliminate aging and to greatly enhance human intellectual, physical and psychological capacities.

The website of World Transhumanism Association (2002) describes Transhumanism as:

> Transhumanism is the study of the ramifications, promises and potential dangers of technologies that will enable us to overcome fundamental human limitations and the related study of the ethical matters involved in developing and using such technologies.

More (2003) defined the term in the broader framework of 'intelligent technology' for 'dynamic optimism' and open liberal society. The use of technology is for the benefit of human beings. This needs to have a clear understanding of the fact that the present world has serious socio-economic divide. So the benefits of technology have to be for all and across the board.

The research in artificial intelligent speech processing systems is directed towards achieving what has been considered as impossible so far. To develop an artificial intelligent speech system that can actively participate in a logical and meaningful human-machine communication act is in line with the objectives and philosophy of Transhumanism. The commercial feasibility of artificial intelligent speech systems in areas like health, education, security, industrial and corporate sector is based on the functional need assessment and user response. Investment opportunities in telecommunication, banking, business management and industries have encouraged products that

are user friendly and purposeful. This is also generating debate in academic circles to discuss the practical issues involved in artificial intelligence speech generation research. The 'speech' ability of an artificial intelligent speech generation system that may 'think' and then 'speak' is critically linked with the 'self-realization' of artificial intelligent speech comprehension and speech generation system (however, there is no evidence in existing literature as yet). Without 'self-realization', artificial intelligent systems cannot have the ability to think and speak and thus socio-cultural flavor of speech may not be possible in a desired logical and meaningful human-machine communication act.

Research Rationale and Focus

This paper looks at the imperatives of artificial intelligent speech production systems having complete realization of languages with socio-cultural imprints. It also generates queries for possibilities and challenges for creating a meaningful, reliable and efficient speech generation system for which there are no logical answers available as yet. Also it attempts to question the intrinsic intricacies involved in professional competence of artificial intelligent speech systems to perform like humans in speech generation and speech comprehension.

Literature Review

It is interesting to note that the research projects, available as credible evidence in literature for developing logical human-machine communication act, have struggled in the last couple of decades to construct artificial intelligent systems to generate meaningful

utterances in speech output despite severe opposition from many people. Some philosophers like Descartes (1991) did not agree with the mechanical explanation of processes involved in human language generation. They believed that human languages are the product of human experience in socio-cultural framework with a realization of 'self' as an identity. This is not possible for machines to have consciousness of 'self' as an individual so human languages cannot be processed by computers. Human language was first seen as a computable act by Alan Turing in his famous Turing Test (1950). Turing's argument was strongly refuted by Searle (1969) who claimed that machines could never know exactly what they were processing as language. With this tough debate on the possibility of computers using human language as artificial intelligent machines or otherwise, a number of speech recognition and generation systems developed in the 70's and 80's (to be discussed later) had serious problems in producing meaningful speech utterances so they did not become popular. The first practical success of speech recognition system was seen in 1990's when the keyboard system was replaced by speech recognition technology for the flight information services provided by the United Airlines of the USA (AI Applications, 2005). In the last two decades, the research in language technology has focused on establishing practical systems which could comprehend natural language and then generate a response in a meaningful speech utterance relevant with the original input of speech utterance generated by human beings.

Research in the area of natural language processing started in the 1940's when the first computers started to appear in the market

(Bott, 1970). Earlier models of machine translations for multiple languages did not find mass recognition as multiple problems appeared across the syntactic and semantic level during translation of languages. Later the interest in natural language processing encouraged the researchers in computer science and engineering to focus on natural language comprehension and natural language generation technologies as the contributory technologies to Natural Language Processing (NLP). The objective of these research endeavors has been to manufacture a system that can comprehend the utterance at semantic and syntactic level for a logical meaning and then to use the appropriate syntax of an utterance as speech output for a logical response. In the last twenty years, research in this area has focused on developing artificial intelligent systems that work according to the syntactic, semantic and pragmatic norms of a human language and which function as per rules of situation and conversation in natural human communication (Barr, 1980, p6). The increasing processing speed of computer systems for large database of linguistic information is surely going to help the artificial intelligent systems to choose the correct words and the correct syntactical structure to convey the correct meaning at the appropriate time and a logical response to speech input to the system.

Limited processing speed and small database were some of the constraints for the earlier speech based processing systems of human language. Complex language structures could not be handled by these systems and their functionality remained focused on limited communication of interrogative and declarative utterances (Barr, p6). PROTOSYNTHEX-1 developed by Simmons (1966) and Semantic

Memory developed by Quillian (1968) are the first models to focus on processing of natural language text and they could easily retrieve information by using indexing techniques. These text based systems could not respond to any query which was not part of the frame of reference in the memory. The data processing of these systems had serious problems with the semantic approach. A successful language processing system needs to have frames and semantic networks for logical and meaningful responses. In language processing programs, grammar is used in parsing to 'pick apart' the sentences in the input to the program to help determine their meaning and thus an appropriate response (Barr, p.7).

William Woods' processing system (1973), LUNAR worked with a program called procedural semantics to get the appropriate response for each input in the system. This program tried to address many of the problems in working with English grammar. SHRDLU developed by Winograd (1975) showed some reasoning ability to have meaningful dialogue with a human being about colors and shapes of blocks. This program influenced the later researchers to develop natural language processing programs based on dialogue in human-machine communication act. The ability for a program to 'converse' is important to its functionality for being 'creative' in speech production. The most pivotal idea of using 'frames' for natural language processing (NLP) was proposed by Minsky (1975). A comprehensive database with frames as prototypes are used for multiple referencing for the analysis of form and manner of dialogue for chronology of situations, simple to complex utterances and simple to complex situations and objects. The later models also relied on

continuous referring to comprehensive database for the output in written text and speech. Bobrow's GUS (1977) and Schank's SAM (1977) are systems for natural language processing which tried to keep maximum number of possibilities of utterances in a given situation and out of that the most appropriate word sequence was selected for the speech utterance as response. Such systems could work very efficiently for a traveler looking for the possibilities of flights or a writer looking for information on a particular situation or script. These systems could not incorporate any 'creative' response for any unfamiliar input which is not programmed or not present as reference in prototype frames.

Speech generation by an artificial intelligent system requires serious deliberation on the part of researchers. This important area moves beyond the working pattern of earlier systems as it requires a 'creative' use of language. The speech comprehension and speech generation capacity of a system is restricted if it has access to a limited database of syntax and lexicon. For an artificial intelligent speech generation system, the competence of language processing needs to incorporate multiplicity in speech for several meanings of the utterance with the syntactic structure with different stress and intonation, and for different syntactic structures for the same meaning. Artificial intelligent systems cannot just mimic the human speech input or generate utterances from the limited choice available in the database. Creativity is a critical aspect for speech generation by artificial intelligent systems. Languages used by humans have a variety of lexicons for different situations which adds to the flavor of the language, for example synonyms. As discussed by Stephen

Wilson (1995) in his article titled *Artificial Intelligent Research as Art*:

> There are understanding computer programs that reduce all mentions of humans consuming solid nourishment (eating) to a primitive internal concept. This strategy allows the programs to proceed with following stories and making inferences about meanings. Humans, however, do not just eat in one way. Sometimes we, gobble, gluttonize, devour, gulp, nibble, sample, gnaw, feast and savor and so on. .....Connotations are as important as denotations.

Shades of meaning and the multiple usage of lexis in different situations is a normal feature of human languages. Each lexicon may give shades of meanings with different stress and intonation pattern in a speech utterance. An effective speech comprehension and speech generation system must have the database and the frames of reference for multiple usages of lexical items in different situations for it to perform as a logical and rational partner in human-machine communication act. Access to a large database of words and syntactical structures makes it possible for artificial intelligent systems to produce meaningful speech utterances as it helps in generating multiple word sequences. The reliability of the speech utterance increases manifold as the utterance becomes relevant to a situation. A lexicon list offers a small number of word choices for speech generation systems like a weather forecast system or flight information system. This system was developed by Zue et al. (2000) and it offers functionality in speech generation in limited situations with restricted choices of lexicon. As compared to this, Varchavskaia et al. (2001) developed a model for speech generation with a capacity to use complex and multiple word choices in speech generation

systems. This model is also used for humanoid robots like Kismet (2006& 2011). Werker et al. (1996) focused on infant speech as the model for speech generation systems. The infant speech is considered to be the starting point for the artificial intelligent speech generation systems as they contain utterances which are short and simple and its lexicon selection and usage are not in isolation (Aslin et al, 1996). Brent & Siskind (2001) developed a model highlighting that single lexical items help in acquisition ability of infants so it can be modeled on systems. Repetition is another feature of infant speech which when adopted by speech generation systems imply that the system is struggling to get the right meaning of an utterance and needs immediate attention for a correct input (Hirschberg et al, 1999). Fitzpatrick (2003, p120) stressed that for an efficient speech recognition system, it is crucial that with the addition of lexical items, the phonetic quality is retained for increase in range of words and comprehension. Fitzpatrick (2003) stressed that this system conforms the phonetic quality of the word to a unique value for future referencing. This is essentially linked with the cultural and social placement of that utterance so any future development in the model requires refinement for socio-cultural relevance and understanding for a meaningful comprehension of speech utterance and then generation of a logical and relevant response in speech to the earlier speech input.

The development of prototypes like Kismet (2006 & 2011) and ASIMO (2007 & 2011) has highlighted the need of serious debate on linguistic issues for efficient working of artificial intelligent speech comprehension and speech generation system.

These humanoid robots have shown credible evidence of successes in speech processing research. They have shown limited understanding of conformity and understanding of the linguistic and paralinguistic features of human communication through speech. The objective of mentioning speech comprehension and speech processing systems is to show their contribution in the overall speech processing research and to establish the fact that multiple questions raised later in this paper need serious attention for developing an effective and efficient artificial intelligent speech processing systems.

The reliability and validity of logical and meaningful communication depend on the relevant speech utterances generated by the artificial intelligent systems showing maximum conformity and understanding of the linguistic and paralinguistic features of human communication through speech. The importance of facial movements, gestures in aid to stress, intonation and tone of speech utterance are critical for logical comprehension of speech utterance and then generating a valid and appropriate response in speech to the relevant speech input.

As a philosophy, Transhumanism focuses on opportunities and potential of technologies. Within the framework of Transhumanism, as mentioned earlier, artificial intelligence research is aiming to develop systems which can strive to decipher some of the fundamental issues raised in this paper. Research inquiry on relevance of speech utterance is based on Transhuman philosophy. Yudkowsky (2004) discussed the implications of artificial intelligent machines speaking like human beings in a natural situation. The literature review confirms the technological advancement in this

research area. Hughes (2004) also discussed *safe* Transhumanism, in which positive features of technology, including artificial intelligence, are seen as integral components of human society.

Research Methodology

This paper is based on descriptive research. It reports the existing facts (Sarma & Misra, 2006) about functionality of speech generation systems and raises intrinsic questions during the critical discussion for developing a logical, reliable and efficient speech generation systems for a sustainable and meaningful human-machine communication act. It explains the ground realities (Chambliss & Schutt, 2009) and documents the historical perspective (Johnson, 2001) which was earlier scattered and not available in a cumulative form. The paper attempts to analyze the available data by 'creating new interpretations in the process' (Noblit and Hare, 1988, page 9)

Critical Discussion

Human beings use languages as tools for communication. These have structures ranging from simple to complex with layers of meaning in particular situations and a complete adherence to the socio-cultural norms of the society. Fodor (1981) suggested that the speech utterances produced by the artificial intelligent machine need to have cultural and social context for the lexical-syntactical patterns in speech utterances. From the initial stages of artificial intelligence research, when it focused on developing systems which could speak and think like human beings, present day research is moving forward in areas of skill acquisition, reasoning, problem solving and developing human-machine communication in speech and writing.

It is critical to note that in any communication act between humans and artificial intelligent machines; 'thinking' is a close associate for the socio-cultural context of the utterances for humans and machines. For a machine to produce meaningful speech utterance, it is essential that the socio-cultural flavor of the language assists the logical reasoning working for the logical and reliable speech generation. The most important challenge for the researchers is to look for the 'social' aspect of the speech utterance. The present literature does not provide evidence for any successes in this area as yet. The historical perspective in the literature review confirms that the research in artificial intelligence focused on modeling of human intelligence as computable patterns. It is a fallacy to connect the computational process with human thinking process just because both are 'processes'. Human thinking process is a combination of understanding of self, social presence and references from experiences. There is no substantial evidence that the same could be possible for an artificial intelligent system. Another challenge for the researchers is to develop a workable model for computers to 'understand' their existence and act in an unpredictable manner. If developed, how would such a system conform to human understanding of gender, nation, tribes, race, economics and global political patterns? These systems need effective checks to be integrated to enable machines to make a decision for producing a speech utterance or otherwise in a given situation. On the whole, the adjustment of a system in a social context is a larger issue than for it to decide when 'to speak or not to speak'. The future research approach needs to be directed towards establishing a realization of 'self' for the machine in a social context. It is critical for a machine to

locate itself in a social framework and use lexicon in a syntactic pattern that is appropriate to the situation and context of the communication act if it is to claim and effective artificial intelligent speech processing system. The social element integrated with speech generation technology would add value to the reliability of speech utterance as it would be logical, relevant to the context and meaningful for the listener.

Human beings react towards a situation and perceive a situation in a social context. The realization of self in human beings is a product of social behavior of individuals. It is yet not known if computer systems can be facilitated with choices of selection of lexical and syntactical patterns and to generate speech utterances with socio-cultural flavor of the language in meaningful speech utterances. The unpredictability and ability of creativity in human life creates multiple opportunities for self learning and self correction. Human languages are product of collective social learning and wisdom. The process of language generation gets influenced by multiple factors, of which perception of the outer world and situation may be only some constituents. Traditions, customs and cultural framework are important features for language generation in human beings. This leads us to the logical question: could machines have these features of language generation without realization of 'self' by machines as identities placed in a social and cultural context and logical reasoning for speech utterances? This seems a difficult task as individual behavior contributes in the development of personality which comprises of style, mood, attitudes, facial features and gestures. These factors contribute in shaping the distinctive personality of an

individual. This is another crucial reference point for researchers in artificial intelligence interested in developing speech comprehension and speech generation systems. The 'individuality' of a speech generation system would be determined by the 'acquired personality', if it would be possible for such machines to have personality, of the artificial intelligent system which is partly evident through creative, meaningful and logical use of language. It is important to ask questions if all of this is possible. The Transhuman approach does not rule out the possibility as it encourages research endeavors that use technology, including artificial intelligence, for creating state of the art speech systems, but essentially demands scientific evidence for the efficiency of such speech generation systems.

There are genuine syntactical and lexical constrains in the models developed for speech generation (discussed earlier in the literature review). The models developed by Hirschberg et al. (1999), Zue et al. (2000) and Varchavskaia et al. (2001) focused on lexicons as units in syntactic patterns for a meaningful and logical speech utterance. These models struggled for efficient performance in fluency and multiple selection of lexicon and syntax for the same meaning or different shades of meaning of speech utterances. The fluency in speech utterances and resolution of ambiguities are some of the major challenges for the future researchers. The clarification of ambiguities in conversation requires repeated referencing to utterances and words closest in meaning in the database. Human speech utterances are comprised of incomplete sentences and repetitions. Now the question arises: how would computer systems respond to such speech input into the system? There is a strong

chance that repetitions, false starts and long pauses in speech utterances would have an impact on the speech comprehension ability and then the logical and relevant speech generation ability of the system. The logical relevance of speech generation with speech input greatly depends on the ability of the speech system to comprehend the speech input and get the meaning as desired by the speaker. This would help in reducing the level of ambiguity for all partners in the communication act through speech. Speech utterances can be understood in more than one different ways.

In a communication act, the job of a listener is very critical as the listener has to get the meaning in the speech utterance. No two speakers share the experience of the language so they do not have the same language. The experience of a human being and the artificial intelligent machine as listener and speaker consists of speech of other individual speaker, each of whom is unique. The ambiguity or misinterpretation in speech utterances is caused by stress and intonation pattern of lexicons, syntax and context. Affirmative and negative sentences may cause ambiguity in speech utterances. There are more chances of misinterpretation in negative sentences. For example in the following sentence:

Ahmad did not eat fish in the market.

The artificial intelligent speech comprehension system may comprehend it differently if the understanding of stress pattern and paralinguistic features is not part of the speech comprehension system. This sentence could mean that Ahmad ate fish at home or he ate something else in the market or he did not eat but had a drink or

he ate fish somewhere else or someone else ate fish in the market. There can be number of possibilities and multiple meanings are possible for this sentence with different stress patterns.

The artificial intelligent speech system is a receiver as listener and it processes the speech input for a meaningful understanding. As listener, the system has to proactively participate in the communication act and at times the role becomes predictive and anticipatory. It is the job of the listener to decode the meaning in speech utterances. Both human and machine as participants in the communication act, may have serious misunderstanding in speech comprehension if the phonological ambiguity of the following kind gets generated during the communication.

For example:

1. Psychotherapist = psycho-therapist
2. New day= nude, eh?

The ability to interpret and correct utterance keeping in mind the context and reference and repeat utterance for clarity in meanings has to be integrated in speech comprehension system as active listener.

An efficient artificial intelligent system must incorporate the understanding of stress and intonation pattern, gestures and facial features, possibility of ambiguity in speech utterances and efficient audio-visual synchronization during the speech act for complete understanding of the meaning of speech utterances. There is no evidence of a successful speech generation system which attempts to completely resolve the issue of ambiguity in the existing literature as yet.

Depending on the language use and context, the integration of features of mood, humor and tone in speech utterances, the artificial intelligent systems can act 'creatively' in a human-machine communication act as these are integral constituents of human communication act. These issues are relevant to speech comprehension technology as well, as systems can only process something for speech utterance when they exactly know what they are asked to say. Danlos' (1987) work is considered as an important contribution in highlighting the linguistic basis of artificial intelligence research in speech generation. She considered the following two decisions as the most important for speech generation.

a. Conceptual Choice, to decide about the sequence of required information, the form and the manner of utterance and
b. Linguistic Choice, selection of syntactical structures and lexicon

This is integral for a speech generation system. The competence of a speech generation system depends on its ability to make logical linguistic choices for structures and lexicon and conceptual choices for level and form of speech utterance. Kismet (2006 & 2011) and ASIMO (2007 & 2011) as speech comprehension and speech generation systems have shown considerable efficiency in achieving this competence but the rate of fluency and selecting an appropriate response for multiple and unpredicted situations and adjusting to the socio-cultural relevance of speech utterances are some of the features which still need attention in the artificial intelligence research in speech processing.

It is anticipated that artificial intelligent systems would perceive human language in a totally different way as compared to humans. They may use and relate to language in a totally different manner, ranging from possible change in syntax to redefining the meanings of lexicons, perhaps not familiar to human beings. Questions still arise: would the cultural and social information carried by languages be of any relevance to these machines? The semantic values of lexicons will surely change when they will used by artificial intelligent machines, for example words like benevolence and chivalry, when the traditional and cultural references are not retained as valid information for speech generation systems. As a linguist, is it practical to think of a situation when intelligent machines would view language from a redefined context and situation from the view point of machine? Perhaps yes, as speech generation systems have shown considerable success in the previous decade and credible evidence is available in the literature review.

The questions raised in this paper aim to improve the performance of speech generation systems to achieve optimum linguistic proficiency in a human-machine and possibly a machine-machine communication act in speech using human languages. There is a serious challenge for artificial intelligence research for creating a speech processing system that is 'aware of itself as an individual machine'. This is not yet achieved even in the latest versions of Kismet (2011) and ASIMO (2011). The survey of the speech systems with some notable successes leads us to the questions: would the future speech generation systems retain the 'ethical' values contained in the language as used by human beings? Would concepts related

with lexical items like gentleness, courtesy, support, sympathy, reconciliation, compassion and leniency retain their existing semantic values or they would be replaced with stringent outcome specific terms relevant to a mechanized way of existence, performance, capacity and efficiency? The answers to these questions are not known as yet in the existing literature.

The discussion on artificial intelligent speech processing systems also leads us to the questions: would intelligent machines be able to produce speech which is creative and also contains the emotions and feelings for powerful expression in rhyme or prose? This would seem a distant reality at this stage and it may take even longer to arrive as we anticipate it today. The language used by humans for the expression of ideas, opinions, knowledge and feelings may one day be used by artificial intelligent machines which would 'own' the language as a communication tool. It is not known as yet, how human languages would be changed in speech if they are to be owned by artificial intelligent systems as languages for communication. It is crucial that these questions get the serious attention of the future researchers if they are deeply interested in effective, efficient and meaningful communicative ability in speech of artificial intelligent machines.

Conclusion

Human-Computer interaction has evolved tremendously in the last two decades from desktop keyboards to complete touch screen systems. The recent developments in natural language technology in speech comprehension and speech generation have made it much

easier for human beings to believe that these systems are more than dumb machines. The visible change in tone of the speech generation system from a monotonous mechanical tone to a more humanoid utterance with intonation and stress pattern just like human beings has also helped in expecting a logical and meaningful speech utterance from a machine. The current trends in different continuing projects of artificial intelligence research demonstrate an interest in developing systems that can generate speech responses to human expressions, gestures, moves and actions (Kismet, 2011 & ASIMO, 2011). This hints towards the beginning of a new era of research for audio-visual synchronization for comprehension of linguistic and paralinguistic features of human communication and then generating appropriate response as speech utterance. The socio-cultural context of gestures and facial features is critical in human-machine communication for audio-visual synchronization during the speech act for speech comprehension and speech generation technologies. The adaptability of speech generation systems to linguistic and paralinguistic features of human languages is decisive for the success rate of such systems in a meaningful human-machine communication act. The market feasibility of such speech generation systems heavily depends on this for a competitive performance in areas like telecom, medicine, law, corporate sector, security, education, tourism and flight services.

It is difficult to expect novelty and creativity from an artificial intelligent machine at this point of time. Languages are unending source of creative combination of syntax and lexicon. The questions like: would speech generation systems be able to produce 'creative'

utterances? Would those be termed creative when they rely on database of language as used by human beings? Would speech generation systems be able to use language just like a native speaker and how the varieties of languages and dialects be incorporated in speech generation systems? These questions are fundamental to the success of artificial intelligence research in speech processing and there may be many more questions from the linguistic point of view in the artificial intelligent research. And for these we do not have the answers as yet, perhaps someday we will.

References

AI Applications, (2005), *Artificial Intelligence Applications for Speech Technologies,* accessed at http://wwwformal.stanford.edu/jmc/whatisai.html on April 5, 2005

ASIMO, (2007), *The Humanoid Robot* accessed at http://world.honda.com/news/2007/c071211Enabling-Multiple-ASIMO-to-Work/ retrieved on December 30, 2007

ASIMO, (2011), *The Humanoid Robot* accessed at http://world.honda.com/news/2007/c071211Enabling-Multiple-ASIMO-to- Work/index.html retrieved on June 21, 2011

Aslin, R., Woodward, J., LaMendola, N. and Bever, T, (1996), Models of Word Segmentation in Fluent Maternal Speech to Infants In *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisitions* edited by Morgan, J. and Demuth, K., New Jersey: Lawrence Erlbaum Associates, pp 450-467

Barr, A., (1980), Natural Language Understanding In *AI Magazine*, 1:1

Bobrow, D.G., Kaplan, R.M., Kay, M., Norman, D.A., Thompson, H. and Winograd, T., (1977), GUS, A Frame-Driven Dialogue System In *Artificial Intelligence*, 8, pp155-173

Bostrom, N. (2003b), *Transhumanist Values*, accessed at http://www.nickbostrom.com/ethics/values.html on March 14, 2005

Bott, M.F., (1970), Computational Linguistics in *New Horizons in Linguistics* edited by J. Lyons, Harmondsworth: Penguin Books, pp 215-228

Brent, M. and Siskind, J., (2001), The Role of Exposure to Isolated Words in Early Vocabulary Development In *Cognition,* 81, pp 33-44

Chambliss, D.F., &Schutt, R.K., (2009), *Making Sense of the Social World: Methods of Investigation*, California: Pine Forge Press

Danlos, L., (1987), *The Linguistic Bases of Text Generation,* Cambridge: Cambridge University Press

Descartes, R., (1991), *Principles of Philosophy,* Dordrecht: Kluwer Academics Publishers

Fitzpatrick, P., (2003), *From First Contact to Close Encounters: A Developmentally Deep Perceptual System for a Humanoid Robot,* PhD Thesis for Department of Electrical Engineering and Computer Science, Cambridge: MIT Press

Fodor, J., (1981). *Representations: Philosophical Essays on the Foundations of Cognitive Science,* MIT Press: Cambridge, MA

Hirschberg, J., Litman, D. and Swerts, M., (1999), Prosodic Cues to Recognition Errors in Proceedings *of the Automatic Speech recognition and Understanding Workshop,* pp 359-374

Hughes, J., (2004), *Citizen Cyborg: Why Democratic Societies must respond to the Redesigned Human of the Future,* Cambridge, MA: Westview Press

Johnson, B., (2001), Toward a New Classification of No-experimental Quantitative Research In *Educational Researcher*, 30:2, pp 3-13

Kismet, (2006), *The Social Humanoid Robot,* accessed on http://www.ai.mit.edu/projects/sociable/kismet.html retrieved on December 29, 2006

Kismet, (2011), The Social Humanoid Robot, accessed on http://www.ai.mit.edu/projects/sociable/expressive-speech.htmlretrieved on June 21, 2011

Minsky, M., (1975), A Framework for Representing Knowledge In *The Psychology of Computer Vision* edited by P. Winston, New York: McGraw- Hill

More, M., (2003), *Principles of Extropy,* version 3.11.2003 accessed at http://www.extrpoy.org/principles.htm retrieved on December 09 2005

Naoko, T., (1993), Neuro Baby in *Siggraph 93 Visual Proceedings,* Art Show Catalogue, ACM: New York

Noblit, G.W. & Hare, R.D. (1988), *Meta-ethnography: Synthesizing qualitative studies,* Newbury Park, CA: Sage

Quillian, M.R., (1968), Semantic Memory In *Semantic Information Processing,* M. Minsky (Ed.), Cambridge: MIT Press, pp 227-270

Sarma, R.P., &Misra, R.N., (2006), *Research Methodology and Analyses*, New Delhi: Discovery Publishing House

Schank, R. and Abelson, R.P., (1977), *Scripts, Plans, Goals and Understanding*, Hillsdale, New Jersey: Lawrence Erlbaum

Searle, J.R., (1969), *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press

Simmons, R.F., Burger, J.F. and Long, R.E (1966), An Approach towards Answering English Questions from Text, *In Proceedings of the AFIPS Fall Joint Computer Conference*, 29, Washington: Spartan Books, pp 349-356

Turing, A.M. (1950), Computing Machinery And Intelligence, In *Mind,* New Series, 59:236, pp 433-460

Varchavskaia, P., Fitzpatrick, P. and Breazeal, C., (2001), Characterizing and Processing Robot directed Speech In *Proceedings of the International IEEE/RSJ Conference on Humanized Robots,* Tokyo

Wilson, Stephen. (1995), Artificial Intelligence Research as Art, In *Constructions of the Mind,* 4:2, seen at http://www.stanford.edu/group/SHR/4-2/text/wilson.html retrieved on March 4, 2006

Winograd, T., (1975), Frame Representations and the Declarative/procedural Controversy In *Representation and Understanding: Studies in Cognitive Science,* D.G. Bobrow and A. Collins (Eds.). New York: Academic Press, pp 185-210

Werker, J., Lloyd, V., Pegg, J. and Polka, L., (1996), Putting the Baby in the Bootstraps: Toward a More Complete Understanding of the Role of the Input in Infant Speech Processing In *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisitions,* Morgan, J. and Demuth, K. (Eds.). New Jersey: Lawrence Erlbaum Associates, pp 427-447

Woods, W.A., (1973), An Experimental Parsing System for Transition Network Grammars In *Natural Language Processing,* R. Rustin (Ed.), New York: Algorithmics Press, pp 111-154

WTA, (2002), *The Transhumanist FAQ: v 1.1,* World Transhumanist Association webpage accessed on http://transhumansim.org/index.php/Transhumanism/FAQ retrieved on March 15, 2005

Yudkowsky, E., (2004), *Collective Volition* accessed on http://www.singinst.org/frindly/colecive-volition.html retrieved on November 04, 2006

Zue, V. and Glass, J., Plifroni, J., Pao, C. and Hazen, T., (2000), Jupiter: A Telephone-based conversation Interface for Weather Information In *IEEE Transactions on Speech and Audio Processing,* 8, pp 100-112