# Emergent Verbal Behaviour in Human-Robot Interaction

Kristiina Jokinen
University of Helsinki
Email: kristiina.jokinen@helsinki.fi

Graham Wilcock
University of Helsinki
Email: graham.wilcock@helsinki.fi

***Abstract*—The paper describes emergent verbal behaviour that arises when speech components are added to a robotics simulator. In the existing simulator the robot performs its activities silently. When speech synthesis is added, the first level of emergent verbal behaviour is that the robot produces spoken monologues giving a stream of simple explanations of its movements. When speech recognition is added, human-robot interaction can be initiated by the human, using voice commands to direct the robot's movements. In addition, cooperative verbal behaviour emerges when the robot modifies its own verbal behaviour in response to being asked by the human to talk less or more. The robotics framework supports different behavioural paradigms, including finite state machines, reinforcement learning and fuzzy decisions. By combining finite state machines with the speech interface, spoken dialogue systems based on state transitions can be implemented. These dialogue systems exemplify emergent verbal behaviour that is robot-initiated: the robot asks appropriate questions in order to achieve the dialogue goal. The paper mentions current work on using Wikipedia as a knowledge base for open-domain dialogues, and suggests promising ideas for topic-tracking and robot-initiated conversational topics.**

## I. Introduction

The paper describes emergent verbal behaviour that arises when speech components are added to a robotics simulator. In the existing simulator the robot performs its activities silently. When speech synthesis is added, the first level of emergent verbal behaviour is that the robot produces spoken monologues giving a stream of simple explanations of its movements. When speech recognition is added, human-robot interaction can be initiated by the human, using voice commands to direct the robot's movements. In addition, cooperative verbal behaviour emerges when the robot modifies its own verbal behaviour in response to being asked by the human to talk less or more.

The robotics framework supports different behavioural paradigms, including finite state machines, reinforcement learning, fuzzy decisions, neural networks and evolutionary algorithms. By combining finite state machines with the speech interface, spoken dialogue systems based on state transitions can be implemented. These dialogue systems exemplify emergent verbal behaviour that is robot-initiated: the robot asks appropriate questions in order to achieve the dialogue goal. When other paradigms such as reinforcement learning and fuzzy decisions are used, further experimental combinations of speech and robot behaviours can be explored. A demo based on this work is described in [10].

The paper is structured as follows. Section II introduces the existing robotics framework and describes silent autonomous behaviours by the robot. Section III introduces relatively simple forms of emergent verbal behaviour, including human-robot interaction initiated by voice commands from the human and human-robot interaction initiated by questions asked by the robot. Section IV discusses current work using Wikipedia as a knowledge base for open-domain conversations. Section V mentions related work and Section VI describes plans for future research.

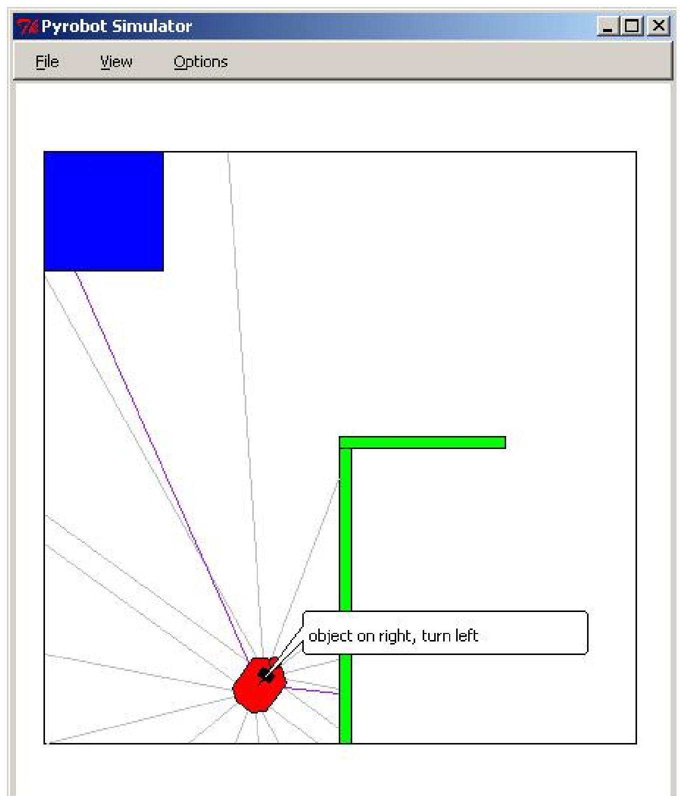


Fig. 1. Pyro Robotics Simulator: The robot "speaks" via a speech bubble

## II. Pyro: Python Robotics

Pyro [1] is an open source Python robotics toolkit for exploring topics in artificial intelligence and robotics, available from http://pyrobotics.org. Although Pyro can be used to control real robots as well as simulations, the paper describes only simulated robots running in simulated worlds. In Figure 1 the robot is a “Red Pioneer” and the world is Tutorial World. The robot has a set of sensors that can detect obstacles and other objects in front, behind, left and right.

The Pyro toolkit does not include speech recognition or speech synthesis, but it does provide a “speech” device that allows messages to appear in a speech bubble as shown in Figure 1. The robot’s sensors have detected the wall on its right, and it decides to turn left.

The robot’s “brain” is a Python program that controls the robot’s behaviour. At the most basic level, the robot can move forward or back and can turn itself to left or right. Appropriate reactions to signals from the sensors can be programmed, so that the robot begins to behave autonomously. Simple autonomous behaviours include Avoid (turning away from detected objects) and Wander (randomly deciding to turn slightly to left or right while moving mainly forward).

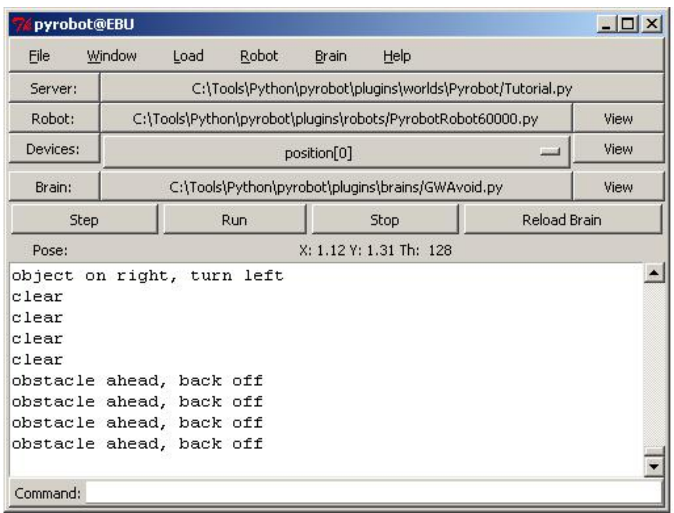


Fig. 2. Pyro Robotics Control Panel

The choice of world type, robot type, and brain type can conveniently be made with the Pyrobot graphical user interface shown in Figure 2. Here a modified version of the Avoid brain has been loaded. This version not only avoids detected objects but also prints brief messages to the console explaining its movements. “Clear” means that the way ahead is clear and the robot simply moves forward. These printed messages provide a basis for the robot’s spoken behaviour in Section III-A.

Commands can also be typed into the command line at the bottom of the control panel. These commands are relatively technical, for example “robot.rotate(-0.3)” to make a small turn to the right. More convenient voice commands are described in Section III-B.

## III. Emergent Verbal Behaviour

This section describes simple emergent verbal behaviour that arises when the speech interface is added to the robotics framework. The system runs on an ordinary Windows laptop. Speech recognition and synthesis are performed by MicroSoft speech engine (Speech SDK 5.1).

The interface between the speech engine and the robotics toolkit is the Python pyspeech package for Windows [4]. This provides convenient functions for text-to-speech and for recognizing words from a given list.

### *A. The robot explains its behaviour*

When speech synthesis is added (not just the speech bubble), messages like those printed in the control panel in Figure 2 can be turned into speech by the text-to-speech component of pyspeech. This is a first level of emergent verbal behaviour. The robot explains its movements while it is performing them. This is spoken monologue behaviour, not dialogue.

However, in this first level of verbal behaviour the messages are produced at every step of the robot’s behaviour cycle. For example, when no avoidance actions are needed the robot just repeats “clear, clear, clear...” constantly while moving forward, which quickly becomes irritating. The robot’s verbal behaviour can easily be improved by saving the last utterance and not repeating the same utterance. Then the robot only speaks when it has something interesting to say.

### *B. Human-initiated verbal interaction*

In human-initiated verbal interaction, phrases such as “go back, turn right” can be recognized by the speech recognition component of pyspeech. These phrases can be used as voice commands for direct human control of the robot at the basic level. However, it would be tedious to specify all the robot’s movements at the basic level. The program in the robot’s brain enables it to perform its own autonomous behaviour, so that human voice commands need to be used only when the robot becomes “stuck”, which happens from time to time.

An example of adaptive verbal behaviour in human robot interaction occurs when the human asks the robot to “talk less” or to “talk more”. When these verbal requests are recognized by the robot, the robot can decide to change its own verbosity level. At high verbosity the robot says “clear, clear, clear” constantly while moving forward. At medium verbosity the robot says new things but does not keep repeating the same thing. If told directly to “shut up” the robot decides to switch immediately to the lowest verbosity level where it continues to think and behave autonomously but keeps its thoughts to itself and says nothing.

### *C. Robot-initiated verbal interaction*

In robot-initiated verbal interaction, the robot takes the initiative by asking appropriate questions in order to find out what the human wants. This kind of system-initiated interaction has been investigated especially in spoken dialogue systems research [6].

The Pyro framework supports several different behavioural paradigms, including finite state machines, reinforcement learning, fuzzy decisions, neural networks and evolutionary algorithms. The intention is to allow experimentation with different approaches to developing robot behaviours. Some uses of finite state machines in robotics are described by the Pyro developers in [1]. When the finite state paradigm is combined with a speech interface, spoken dialogue systems based on state transitions can be implemented in Pyro.

As an example, we implemented a "classical" flight reservation system within the Pyro framework. Dialogue control is managed by the robot's brain program using state transitions. The robot initiates verbal interaction with the goal of finding out what the human wants. In addition to the usual states for finding out the departure and destination places and the departure and return days, the example includes states for summarizing the whole trip and for starting over in case of misunderstandings.

As the purpose is to show the finite state dialogue control as clearly as possible, the robot does not move during the dialogue. The robot's emergent behaviour in this case is purely verbal. Like a human, it stands still and holds a conversation until the purpose of the conversation is achieved.

## IV. Towards Conversational Interaction

The "classical" type of spoken dialogue system described in Section III-C is typically restricted to a fixed domain such as flight reservations. Information about specific flights and seat availability is constantly updated in the database, but the range of concepts that can be discussed (such as destination cities and departure dates) is fixed. In order to extend the range of topics that can be discussed, the single domain database needs to be replaced by an open domain internet-based knowledge source such as Wikipedia. We are currently working on how to use such knowledge sources in spoken dialogues.

### A. Human-initiated conversational interaction

Wikipedia provides interfaces for retrieving articles on a given topic. If the human introduces "Shakespeare" as a new dialogue topic, the robot can obtain the Wikipedia article about Shakespeare. As a first attempt at conversational behaviour, the robot could simply read the whole Wikipedia article out aloud from start to finish. However, this would be a monologue, not a dialogue. It would quickly become irritating in the same way as the robot describing its movements at every step of its behaviour cycle (Section III-A).

As a next level of conversational behaviour, the robot can utilize the paragraph structure of the Wikipedia articles. The human authors have divided the articles into paragraphs, based partly on the idea of one subtopic per paragraph and partly on the idea of a paragraph being a convenient size for reading in one "chunk". The robot can pause after reading each paragraph and ask the human if it should continue or stop.

However, the challenge is how to discuss Shakespeare in an interactive dialogue with the human. Our approach is based on the Topic-NewInfo distinction which we have previously used in dialogue response generation [7] and constructive dialogue modelling [5].

When the Wikipedia article is retrieved, Shakespeare is the Topic that the human has initiated. When the robot selects something from the article to say about him, that will be a piece of NewInfo that is novel to the user and which the human may continue with on her next dialogue turn. Once one of the NewInfos has been selected as the new Topic, the procedure continues in the same way.

From the dialogue management point of view, the robot always has a set of concepts that it expects the user to continue the topic with, if the user is interested in the introduced topic. If the user is not interested in the topic, the user says something else which gains importance as a new element in the context, and the robot's attention is directed to this new topic.

### B. Robot-initiated conversational interaction

When the robot reads out a paragraph and pauses to check if the human wishes to continue, the human may be undecided. In that case the robot can use the structure of the article to suggest different subtopics that might be of interest. Using the section headings of the Shakespeare article, the robot can ask "Would you like to know about his life, his plays, his poems, his style, or his influence?".

In addition to suggesting subtopics after the human has initiated the main topic, it is desirable that the robot should be able to propose entirely new topics. An attractive advantage of Wikipedia is that robot-initiated topics can be selected from its "On this day ..." and "Did you know ..." sections. These topics are potentially interesting and are suitable for inviting a silent human to hold a social conversation with the robot.

The human may reply only "Really?" but that is sufficient to implicitly invite the robot to talk more about the started topic. The human may also pick up on some relevant concept in the robot's presentation, and using this as a keyword, invite more information about the particular concept. The robot can use the Wikipedia article's hyperlinked concepts as expected utterance topics that the user is likely to pick up as the next interesting topic. The robot can thus anticipate the course of the conversation.

## V. Related Work

In ethology, the object of research is the behaviour that is specific to a species. Considering human behaviour, one of the main interests is the social communicative behaviour that is realised through natural language. The development of human behaviour is commonly assumed to be based on temporal repeated patterns, and human communicative behaviour is no exception: we learn languages and conversational strategies by interacting with our environment and fellow humans. For instance, Duncan and Fiske [2] noticed certain interaction sequences between children's and adult's interactions, while Magnusson [9] argues for the so called T-patterns which are behavioural time patterns detected through their order and relative timing, and forming hierarchical structures.

Fong et al. [3] listed several characteristics of human social behaviour:

- express and/or perceive emotions
- communicate with high-level dialogue
- learn/recognize models of other agents
- establish/maintain social relationships

- use natural cues (gaze, gestures, etc.)
- exhibit distinctive personality and character
- learn/develop social competencies

In conversations different event-types recur in the same order, and the whole system of communication is a complex system of relationships between the events. Kendon [8] talks about general principles which "are discoverable and generally applicable, even though the course of any specific encounter is unique". We have claimed [5] that conversational activity is governed by certain communicative principles which have their basis in human coordinated action and in agents' basic willingness to cooperate in order to achieve their underlying goals and intentions.

## VI. Future Work

The agents must establish a common ground [5], and the way this is achieved in conversations is to give suitable feedback. Feedback is needed to tell the partner if their presentation has been understood and whether the agent agrees with the content of the presented information. However, in addition to this kind of feedback the interlocutors also exhibit their personality and character, and give natural cues that indicate to the partner whether they are interested in the topic or not.

The participant's reaction, in terms of showing interest or lack of interest, has not been the focus of much attention in conversational studies except indirectly. The agent's interest in the presented issues has not been explicitly studied. However, in our robot world it is important that the robot shows its interest clearly to the human, and the robot also needs to learn to observe when its human partner is giving non-verbal signals that show interest or lack of interest in the presented issues.

For instance, the human can say "Really?" meaning "I don't believe it" or "I haven't heard about it", eliciting more information and thus encouraging the robot to continue, but this same utterance with different intonation or different body language could also mean "I don't think what you say is true and I don't want you to continue". The detection of non-verbal feedback concerning the partner's interest is thus an important aspect in social interaction and something that both humans and robots in our robot world need to learn.

The social robot needs to learn to distinguish between two alternatives: either the partner's reaction shows interest and allows the robot to continue on the current topic, or the partner is not interested in the topic and the robot will do better to find some other topic to talk about. Awareness of these two contrasting reactions needs to be more or less continuous.

Another pertinent topic for future analysis is the use of non-verbal communication strategies in different cultural contexts. We hope to link this kind of work to research on robot agents as the models can be tested and verified experimentally.

Models of feedback-giving processes are relevant in order to design and develop more natural interactive systems. A major question is: what kind of robot is good for human interaction? Following the path of ethological engineering we can say that robots should have their own function in an ecological sense and they should thus represent a novel "species" (or a novel communicative "function"). The robot can be a behavioural mosaic and the novel functions will determine the best configuration for how the robot's development should take place, i.e. the robot should be able to develop and learn its environment. On this view the evolution of human robot interaction may go in a direction that is not directly predicatable on the basis of the current more human-centered views of intelligent agents.

Such novel communicative skills cannot be achieved only by improving interaction technology but also by advancing the systems communicative capability [5], i.e. by modelling the users natural dialogue strategies and integrating the models in the systems behavioural component. Dialogue strategies such as feedback of accepted requests, acknowledgement of understood commands, management of turn taking etc. are often coordinated using non-verbal signals, and the systems should thus accept a wide variety of modalities in which the users convey meanings and control the interaction: they should be capable of sensing and interpreting communicative signals expressed by gestures, facial expressions, eye-gazing, and body posture.

## References


[1] D. Blank, D. Kumar, L. Meeden, and H. Yanco, "The Pyro toolkit for AI and robotics," *AI Magazine*, vol. 27, no. 1, pp. 39–50, 2006.

[2] S. Duncan and D. W. Fiske, *Face-to-face interaction: Research, methods and theory*. Hillsdale NJ: Lawrence Erlbaum, 1977.

[3] T. Fong, I. Nourbaksh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and Autonomous Systems*, vol. 42, pp. 143–166, 2003.

[4] M. Gundlach, "pyspeech: Python speech recognition and text-to-speech module for Windows," 2011, http://code.google.com/p/pyspeech/.

[5] K. Jokinen, *Constructive Dialogue Modelling: Speech Interaction and Rational Agents*. Chichester, UK: John Wiley & Sons, 2009.

[6] K. Jokinen and M. McTear, *Spoken Dialogue Systems*. Morgan & Claypool, 2009.

[7] K. Jokinen and G. Wilcock, "Adaptivity and response generation in a spoken dialogue system," in *Current and New Directions in Discourse and Dialogue (Text, Speech and Language Technology, Volume 22)*. Kluwer Academic Publishers, 2003, pp. 231–234.

[8] A. Kendon, *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge University Press, 1990.

[9] M. S. Magnusson, "Discovering hidden time patterns in behavior: T-patterns and their detection," *Behavior, Research Methods, Instruments & Computers*, vol. 32, pp. 93–110, 2000.

[10] G. Wilcock and K. Jokinen, "Adding speech to a robotics simulator," in *Proceedings of 3rd International Workshop on Spoken Dialogue Systems (IWSDS 2011)*, Granada, To appear.