Investigating spoken dialogue to support manufacturing processes

Professor Rob Gaizauskas
University of Sheffield
Executive Summary

We report on the Connected Everything: Industrial Systems in the Digital Age project “Investigating spoken dialogue to support manufacturing processes”. We review the research challenge addressed by the project, discuss the approach we took to addressing the challenge, describe how we implemented that approach, report our results and discuss wider applications and future work.

We set out to explore the potential for spoken language interfaces in manufacturing and to assess whether current automatic speech recognition (ASR) technologies were mature enough to be deployed in manufacturing settings. Spoken language dialogue is appealing in manufacturing settings because it is hands-free, natural and highly flexible. As digital information systems come increasingly to underpin manufacturing operations, access to these systems will be required from the shop floor and spoken language dialogue could be the ideal medium to interface between the workforce and these systems.

We addressed this challenge via a two-pronged approach of: (1) visiting and interviewing manufacturing partners in both hi-tech (Factory 2050, AMRC, Sheffield) and more traditional manufacturing settings (ELG Utica Alloys Sheffield) to investigate potential use cases and make audio recordings in manufacturing settings with the existing workforce in order to assess the viability of current ASR, and (2) designing task settings in which to collect and study human-human dialogues that simulate scenarios reflecting manufacturing use cases in which spoken language human-computer interaction would be valuable, specifically a lifting and positioning scenario where a robot co-worker holds and repositions a work piece under voice control as a human works on it and an intelligent decision support scenario where an intelligent agent offers advice to a human worker engaged in repairing or servicing a complex artefact, such as an internal combustion engine.

Outcomes of the research, in brief, are: (1) we established that interest in and the potential for spoken language interfaces in manufacturing is high, both in hi-tech and traditional manufacturing settings, with maintenance, repair and overhaul (MRO), complex assembly and shop floor data entry and querying being promising applications areas right now; (2) we established that current ASR techniques are mature enough to be deployed in real manufacturing settings; (3) we produced a collaborative robotic demonstrator utilising a speech interface to control movement, which has been displayed at several events, and highlighted the need for natural dialogue interfaces; (4) we determined that a key challenge for spoken language systems to support workers in MRO is the acquisition of knowledge and data to underpin these systems, which may be held in non-digitised (e.g. paper) and/or free text (e.g. instruction manuals) form and may require innovative solutions to acquire at scale; (5) we designed two experimental scenarios and associated experimental protocols for capturing human-human task-based dialogues that will shed light on the detailed requirements for human-computer dialogues in manufacturing tasks; (6) we wrote and submitted three grant applications based on work carried out in this project, two of which were highly ranked but unsuccessful and are currently being prepared for resubmission and a third which was successful and is now exploring prospects for commercialising insights gained from this research; (7) the researchers gained a significantly richer and deeper understanding of this cross-disciplinary area, which they will bring to further activities in this area.

Future work planned and already underway includes further grant applications in this area, execution of the experiments designed in this project, further development of speech interfaces for cobots, and the development of a nascent University-sponsored commercial enterprise in this area.
1. Research challenge

Digital technologies are bringing about revolutionary changes in manufacturing. Amongst the many emerging capabilities envisaged to form part of this revolution are: (1) human-robot coworking, where the complementary strengths of humans and robots in industrial settings are exploited to deliver performance superior to that achievable by either on their own, and (2) intelligent decision support, where human factory workers, no longer simply machine operators, are supported by rich information systems in making decisions. For example, imagine a human and robot working together to disassemble and recover materials from a spent jet engine. This scenario requires situated interaction, where both parties refer to objects in their shared setting, the human perhaps instructing the robot to use its gripper to remove a specific part; in deciding how to proceed, the human might wish to consult external information sources to determine, e.g., whether removing the part might release a hazardous chemical. Developing these capabilities and integrating them into the work environment is a key part of the Industry 4.0/“factory of the future” vision.

Core to both human-robot coworking and intelligent decision support is communication between humans and machines, whether the machines be robots or information systems. Human-machine communication may take many forms and it is not straightforward to determine, in a given setting, which form human-machine communication should take. Our project chose to investigate the potential for spoken natural language dialogue systems in manufacturing environments.

Our research hypothesis was that spoken natural language dialogue has the potential to be uniquely effective and enabling as means of communication between humans and machines in manufacturing environments, specifically for human-robot coworking and for decision support, and that it is now a mature enough technology to be exploited in these environments.

Why Spoken Natural Language Dialogue? Spoken natural language dialogue has a number of attractive features for the manufacturing environment, including: (1) it supports “hands-free” and “eyes-free” working, allowing workers to carry on with physical tasks while interacting with a machine (robot or information system); (2) it is “natural” for the human participant, meaning little or no training may be required to use it (3) it is highly flexible, allowing mixed initiative interaction between human and computer, communication at different and varying levels of detail and the capacity to switch rapidly between multiple task contexts. Furthermore, speech recognition technology has now reached the point of viability and user acceptance for a whole range of potential applications, as consumer products like Amazon’s Alexa and Apple’s Siri demonstrate.

The project aimed to produce an evidence-based assessment of the potential for spoken dialogue systems (SDS) in human-robot co-working and intelligent decision support in real manufacturing environments, identifying specific scenarios where SDS would be useful, determining requirements on those systems, assessing the extent to which existing technologies may be sufficient and where research challenges may lie, and laying the groundwork for future work in this area.

2. Approach

We followed a two-pronged approach. One stream of work set out to investigate manufacturing settings where SDS might play a role and, by observing current practice, to identify use cases for SDS that would offer improvements to current practice. We aimed to carry out multiple visits to factories, in two different settings (hi-tech vs traditional), carry out informal interviews with employees and observe examples of current working practice, making recordings (video and/or audio) and manual notes. We proposed to analyse data gained through these visits to conceive a number of scenarios for dialogue-based interaction in the different settings, determining tasks that could be supported by dialogue technologies and identifying requirements and challenges for future technology development. We also proposed to investigate the robustness of automatic speech recognition technology in relevant factory settings, considering both cloud-based ASR (e.g. the Google speech API) and state-of-the-art open source platforms.

The second stream of work proposed to design and carry out experiments for collecting human-human task-oriented dialogues for scenarios identified in the first stream. We aimed to recruit participants and prepare robotic systems/props for use as role players and then record examples (video and audio) of dialogue-based interaction for different tasks. We planned to analyse this
data to determine the requirements on dialogue technologies and assess whether or how close state-of-the-art dialogue system capabilities are to being able to realize these requirements. We also aimed to use selected recordings to make a number of demonstration videos for dissemination of our ideas and to assist in gathering feedback and further requirements in subsequent factory visits. While there has been prior work on dialogues for team and collaborative working in task contexts in various domains, to the best of our knowledge no studies of use cases and requirements for dialogue in real world manufacturing settings exist. Work in other domains cannot be assumed to carry over to the manufacturing domain, either in terms of the functional requirements that the task settings place on the SDS or in terms of other aspects of the environment, such as noise, safety of workers, or willingness of the workforce to participate. By contrast, other researchers are investigating human-robot collaboration in industrial contexts, including manufacturing, but are not investigating spoken dialogue. Thus our project will be the first academic study to investigate the potential for SDS in real manufacturing settings.

3. Implementation

Stream 1: Observing Current Practice in Manufacturing

The following activities were carried out as part of investigating real manufacturing settings where SDS could play a role:

Ethics. Because we intended to interview and make sound and potentially video recordings of staff in manufacturing settings we needed to make sure we following best ethical research practice. An ethics application was prepared and submitted to the University of Sheffield Research Ethics committee and consent obtained for the research.

Site Visits. We paid multiple visits to the sites of our two partners, who are engaged in very different manufacturing enterprises:

ELG Utica Alloys, a revert management company specialising in recovering high value alloys from machinery that has reached the end of its lifetime (e.g. jet engines). They are a good example of a traditional manufacturing enterprise: large investment in plant and equipment; relatively little
digitisation, particularly on the factory floor; noisy; dirty; lots of mechanical and electro-chemical processing. We visited two of their facilities in the South Yorkshire area, one concerned with receiving a new spent jet engine and carrying out the initial breaking apart and also housing facilities for more general alloy reclamation/processing; the other concerned with sorting, identifying and storing alloys and producing new bespoke alloys from them. In the visits we went on detailed tours of the facilities and staff provided an overview of key activities and processing.

AMRC’s “Factory 2050”, which is developing numerous advanced digital technology prototypes for a range of external business collaborators, is an example of an organisation focused on the new wave of advanced digital solutions for manufacturing. In our visits we discussed and were given demonstrations of a number of their prototype development projects. In particular we were shown examples of digital work instructions they were developing, using both augmented reality headsets and video screens and integrated with sensing equipment that could verify aspects of an assembly process; we were also shown examples of robotic arm assembly and shaping and discussed maintenance, repair and overhaul (MRO) processes, particularly with respect to the aircraft industry and how spoken language technologies could help with the extensive, labour intensive processes of inspecting and repairing aircraft.

Use Case Development. Based on the site visits we identified and refined several use cases where it appeared spoken language technology could be most useful in manufacturing.

a. Lifting and positioning. Stemming from ELG Utica’s requirement to have flexible mechanical support during the initial stages of breaking down a large, potentially multi-ton, jet engine, we explored the scenario of a robotic arm holding a jet engine and positioning it, under voice control, to enable a worker to access and remove portions of it. This process is currently carried out by multiple people, coordinating and utilising fork-lifts to position the engines. Automation of this process would reduce the number of workers required, but would require workers to be able to direct the robot using a hands-free interface. During visits to the factory we interviewed different staff (executives and workers) to explore the idea and feasibility of a lifting and positioning robot with voice control in more detail and received generally positive feedback.
b. **MRO Support.** After initial discussions with the AMRC, we reviewed the current inspection practice of one of AMRC’s partners in the aerospace industry and examined many maintenance manuals and videos from different areas of manufacturing, to explore how an interactive spoken language system could be of help in carrying out repairs and maintenance activities, where a worker may need first to gain access to a particular component within a complex assembly and then needs to check or set some particular parameter (e.g. a valve clearance; a rivet diameter). We investigated the type and format of information that is created, stored and used in MRO, especially in aerospace, and to some extent in the automotive domain. And we looked at the current systems for managing and accessing such information.

c. **Data entry and querying during process execution.** In reviewing ELG Utica’s current practices and discussing them with management we discovered a simple but potential highly useful application in supporting voice drive data entry and querying on the shop floor. Using voice to directly enter data into, or query data from, their enterprise resource planning system would enable them to replace their current highly inefficient process of recording weights and process settings on paper, and then aggregate cards and walking them to a remote computer for subsequent data entry or going to a remote computer to query settings pertaining to a similar previous process. To help us to understand where we might best situate such voice interface technology in a factory, we carried out further tours of the factory floor and discussed the activities carried out at the various different locations and in relation to different plant/machinery. In discussions with employees during tours, we also identified various types of information tasks that are currently carried out in relation the key processes and various locations in the factory, and we gathered examples of actions that employees thought would be useful if mobile, hands free voice access to their information system were available (eg querys about the location of a product).

**Audio Recording.** At ELG Utica we collected audio recordings at key locations within one of their (very noisy) factories. These were collected according to a protocol we developed to ensure comparability across locations. Specifically we designed a simple dialogue based on use case 3 (c) above and got an ELG Utica employee to dictate the data to be entered in response to a prompt given by one of our researchers. Having recorded this dialogue in different positions and noise levels around the plant, we took the data and ran it through two different automatic speech recognition (ASR) systems: the Google cloud ASR service and a version of the Kaldi open source ASR system, specially tuned by us for the task at hand.

**Robot speech interface.** To explore the use case mentioned in 3 (a) above, we implemented a speech interface on a collaborative robot in our labs, which would enable users to direct the movement of the robot to pick and place items using directional commands. The mapping of commands directly to motor actions resulted in an interface that was awkward to use without practice. This highlighted the need for participatory experiments to identify the types of commands and actions that would be intuitive to non-expert human users.

**Stream 2: Gathering Simulated Human-Computer Interaction Dialogues for Manufacturing.**

**Ethics.** Because we intended to involve human participants in our simulated human-computer interaction dialogues and make sound and potentially video recordings of them, we again needed to ensure we carried out our research according to best ethical research practice. A second ethics application was prepared and submitted to the University Research Ethics committee and consent obtained for the research.

**Scenario Specification.** Based on our Use Case Development work in Stream 1 and considerable wider reading and discussion with academic colleagues with expertise in engineering and automotive maintenance, we chose two scenarios for our simulated human-computer interaction dialogues. Both centred on tasks involving a model car engine, the Haynes V8 Model Combustion Engine:

**Experiment 1: “Lifting/positioning support for Engine Maintenance, Repair and Overhaul”**

This experiment involves two human participants, one playing the role of an automotive engineer who is charged with replacing a piston in the model car engine, the other playing the role of a co-worker whose job is to lift and position the engine as instructed by the engineer, so that the engineer can carry out his work on it. The aim of the setup is to encourage role-players to talk naturally and generate utterances that communicate information
about spatial positioning/alignment of the engine. In the experiment the engine is fixed to a bracket on a robotic arm (the Kuka iiwa robot arm located in the Sheffield Robotics lab) and the lifting and positioning is done by a human operating the robotic arm (played by someone in Sheffield robotics with considerable training, experience and expertise in operating the arm). The other worker who carries out the piston replacement is a recruited participant. This participant has the task explained to him/her in advance and is given materials to read before the experiment that explain what needs to be done to carry out the task. To enable this experiment, the model engine was modified, and a bespoke attachment fabricated, to enable it to be mounted on a robot arm in a way that would enable it to be oriented through multiple planes and dismantled/rebuilt.

Experiment 2: "Intelligent Decision support for Engine Maintenance, Repair and Overhaul".

This experiment again involves two human participants. The task is to inspect, record and, if necessary, adjust the valve clearance in one of the cylinders in the model engine, a task commonly carried out as part of standard engine maintenance. Carrying out the task involves disassembling the engine to expose the parts that need inspecting and adjusting and then reassembling it. One human plays the role of the engineer carrying out the task; the other plays the role of an expert who understands how to disassemble/reassemble the engine – specifically which parts need to be removed and in what order in order to enable the task to carried out – and who knows what the valve clearance ought to be. The expert's role is to help the engineer in the task by providing information on request. All communication between the engineer and the expert is via spoken language. The expert is not able to see the engine and the engineer has to describe to the expert what he sees or what problems he has. Once again, the aim of the setup is to encourage role-players to talk naturally and generate utterances that communicate information that is necessary/helpful to someone carrying out an engine MRO task. To play the role of the expert a participant with some knowledge of engine maintenance is required. They are given additional materials about the specific model engine in advance of the experiment and have materials at hand to consult during the experiment. To play the role of the engineer we also recruit participants with some knowledge of engines and give them some background material on the task before we start.

Data Gathering. For both experiments participants are observed undertaking the simulated tasks, and video and audio recordings made as they do so. These sessions include a few questions before the scenario (e.g. how much experience they have of internal combustion engines and engine maintenance/repair) and afterwards (e.g. why they did what they did during the scenario). The intention was to transcribe the recordings as/when appropriate and depending on available resource.

Experimental Protocol Development. Once the experimental scenarios were identified (point 2. above), precise protocols for the experiments needed to be designed. This included: specifying the questions to be asked pre- and post-session; compiling and editing background materials to be given to the experimental participants for reading before and consulting during the experiments (this involved extracting relevant information from instruction manuals); precisely specifying in writing the tasks the participants are to carry out (this involved gathering a number of examples of similar task instructions from the web, identifying a set of generic task steps for 'valve clearance' inspection and adjustment, and adding specific details in relation to the model engine); deciding the lab space to use for experiments, positioning of video cameras and microphones, etc.

4. Summary of Results

Overall, the results of the work have been very encouraging, though some of the objectives remain to be completed. Here we first list the major outcomes of the work so far and then list some of the lessons learned from the work, which may be instructive for further research work in this area.

Outcomes of the work.

Interest in and the potential for spoken language interfaces in manufacturing is high, both in traditional manufacturing settings and in newer, Industry 4.0-aware settings. For traditional manufacturing, investment in robotics may be some way down the road but spoken language access to databases on the shop floor would be valuable right now. In more digitally advanced environments the most compelling and immediate potential applications lie in intelligent decision support and data capture for maintenance, repair and overhaul (MRO) and for intelligent decision support in complex assembly tasks, potentially in conjunction with visual digital work instructions.
Existing ASR techniques appear to be robust enough to use in practice in real manufacturing settings. Both the Kaldi and Google cloud-based ASR services yielded usable results in terms of recognition accuracy. Kaldi is more tailor able to specific tasks than the Google ASR service and has the advantage of not requiring internet access or internet latencies to run; however, it has lower general vocabulary accuracy. Both would need some adaptation for each specific use case.

We have created a demonstrator of human-robot coworking using a speech interface to direct the actions of the robot. Initial tests by the project team indicate that mapping vocabulary to motor actions (e.g. “Move right arm up”, “Move left arm out”, “open gripper”) are too prescriptive, and that more abstract dialogue (e.g. “more/less” “back a bit”, “towards me”), and primitives (e.g. “pass the spanner”) would be more user friendly.

A key finding was that the current state of information systems and knowledge used in MRO presents an important challenge and constraint on future development of spoken language systems for intelligent decision support or data capture. Much task knowledge is tacit (in the minds of experts) or embedded in text-based manuals (often non-digital); information systems in current practice are often not in digital form; for many complex pieces of equipment, e.g. aircraft, there are issues to do with data sharing and common data formats, between OEMs and organisations servicing and operating the equipment. While there are on-going efforts to digitise information and provide technology solutions for resource sharing, new voice interface applications will be limited to providing access to systems available in digital form. Industries will need to be persuaded of the benefits of what can be costly digitisation processes. Furthermore innovative methods are required to automatically extract the information needed for spoken language systems in MRO from available sources. One such method would be to apply information extraction techniques to obtain task models from manuals and “how to” videos on the web.

We have designed two novel experimental scenarios for capturing simulated human-computer dialogues, based on real manufacturing scenarios, one for human-robot interaction where a robot is lifting and positioning an artefact under voice control to support a human in carrying out maintenance work on it, the other for human-intelligent agent interaction where a human needs advice/instructions on how to carry out a maintenance/servicing task. Detailed protocols for these experiments have been specified, but at the time of writing the experiments have not been run (see Future Work below).

During the course of the work, three related grant applications stimulated by this project have been submitted. The first was for the EPRSC Bright Ideas call in August 2017 and was on the topic of learning task models for robots from how-to documents and videos on the web. It was short-listed for interview in a very highly competitive round, but not awarded. The second was for a University of Sheffield internal Knowledge Exchange fund. Its aim was to work with ELG Utica, a partner in the current project, to build a demonstrator to support the “Data Entry and Querying” use case described above under Implementation, Stream 1 3(c), with a view to that being a first step towards commercialising the ideas. This was initially recommended for funding, but then not supported when a budget cut to the available funds was imposed. The third was also an application to a University of Sheffield internal fund, this time for initial commercialisation work to explore the same scenario as the previous application. This was supported and work towards starting up a business to develop technology to support the Data Entry and Querying scenario is underway (see Future Work below). An MSc project student is currently building on the robotic demonstrator work, by creating an interface for another robot, which will use a more natural language interface.

The cross-disciplinary understanding of the researchers has increased immeasurably, placing them in a position to take work in this area forward, via both further research projects and efforts to build a commercial enterprise to exploit this understanding (see Further Work below).

**Lessons learned.**

Getting ethics right in this sort of research is difficult and highly time consuming. Approximately one third of the time of one of the two part-time researchers on this grant was spent preparing and submitting the ethics applications for the work, albeit some of this overlapped with detailed specification of the experiments to be carried out. One of the problems here was the very prescriptive University Ethics forms and procedure, which is not really suited for this sort of exploratory research.

Designing appropriate scenarios for simulated human-computer interaction dialogues, fully specifying protocols for human participants in such experiments and then recruiting participants is very
time-consuming. It is also difficult in a University environment where the pool of experimental participants (students in the appropriate disciplines) is large but continually moving in and out of availability due to various term and holiday constraints. More generally the sort of cross-disciplinary understanding, in often very technical subject areas, that a project such as this requires is extremely difficult to obtain and then capitalise upon within the time frame of a feasibility study.

Talking to real users (Stream 1 in Implementation above) is extremely valuable as it may throw up scenarios not conceived of by the researchers in advance. Project partners had originally indicated human-robot interaction would be of significant interest in practice and second to that human-intelligent agent interaction for decision support. In fact, as the project progressed, the most appealing application was discovered to be a straightforward data entry application, with the potential to save considerable time in capturing data from the shop floor and improve its quality. This is not an application we had foreseen at all. Following that, human-intelligent agent spoken language interaction for decision support appears to be the next most appealing application, while human-robot interaction using spoken language was viewed as interesting but some way off, largely because of the lack of penetration of robots in the industries we looked at. Thus our preconceived priorities needed to be reversed and something we had not considered came to the fore.

5. Wider applications

We set out to investigate the potential for spoken language interaction in manufacturing settings, specifically to investigate whether it could prove useful for human-robot interaction and intelligent decision support. Our efforts and findings related to this investigation are detailed above. In addition to this we can make two observations related to wider application.

First, through our wider contacts we became aware during the project that the idea of spoken language interaction for hands-free intelligent decision support would also be of interest in the construction industry. In particular the rise of building industry standards, like BIM (building information management), which are meant to support all stages of the building life cycle from design through construction to post-occupancy maintenance, means there will be a requirement for access to information systems by several groups of workers whose work demands use of their hands. Specifically during construction and during maintenance workers are engaged in manual tasks, but may need to consult digital information about where building elements are or are meant to be located and to record work steps they have carried out. As is the case for manual workers in manufacturing, hands-free access to the requisite digital information sources could offer many advantages. We have met with members of our School of Architecture to discuss this possibility and are engaged in setting up meetings with a major architectural firm to discuss these ideas.

Our second observation is that there is a wave of rising interest in the application of augmented reality (AR) headsets in manufacturing, with much discussion of "the augmented worker". However, while the motivation for use of these headsets is to provide hands-free information to workers, the mode of interaction with the headsets is still predominantly manual or involves very limited voice commands. Adding dialogue-based spoken language interaction capabilities to AR headsets is likely to increase the utility of these headsets and the ease of interaction with them. Furthermore, allowing ASR output to be confirmed visually via AR headset is likely to improve the quality of spoken language interaction. Thus combining spoken language dialogue with AR display looks to be a very fruitful way forward.

6. Future Plans

We would like to highlight three aspects of our future plans.

First, as indicated above we are seeking to commercialise ideas that have arisen from this research with support from the University of Sheffield commercialisation team. At present we are conducting market research and seeking partners for a pilot of the technology, as well as exploring investment options.

Second, we are planning to run the two experiments we have designed during this research. We have recently had agreement from partner AMRC that we can recruit participants from amongst their apprenticeship trainees, which should be a better solution than recruiting from amongst the university student population.

Finally we are planning at least two grant submissions in this area. One is for the open EPSRC Adventurous Research in Manufacturing
call, and will be in the area of building a demonstrator of spoken language dialogue combined with AR headset display for an application in MRO. The other will be a resubmission of the Bright Ideas proposal described above (see Results) as an EPSRC responsive mode proposal to explore the idea of mining task models from equipment manuals and how-to documents to support spoken language dialogues for intelligent decision support in manufacturing.

7. Conclusions

In conclusion, our project has established both the viability and demand for spoken language interfaces in manufacturing and in related areas such as building construction and maintenance. In manufacturing intelligent decision support in areas such as complex assembly and maintenance, repair and overall appear to be the most promising areas. However, more straightforward applications such as data entry and querying from the shop floor during routine process execution also have significant immediate potential and do not require significant new technology development. Designing and gathering task-based human-human dialogues to simulate the sort of human-computer dialogue-based interaction that more advanced applications such as human-robot coworking and intelligent decision support, is highly valuable but demanding work. We have made significant steps in this direction by conceiving and specifying experimental protocols for two exemplar tasks and plan to complete these experiments as soon as possible.

8. Feasibility study team members

The study was conducted by a team of researchers from the University of Sheffield:

Professor Robert Gaizauskas, Computer Science
Dr Emma Barker, Computer Science
Dr James Law, Robotics
Dr Samuel Fernandez, Postdoctoral researcher