Using machine learning to emulate human hearing for predictive maintenance of equipment

Dinesh Verma*[a], Graham Bent[b]

[a]IBM TJ Watson Research Center, 1110 Kitchawan Road, Yorktown Heights, NY 10598, USA
[b]IBM Emerging Technology Services, Hursley Park, Hants, UK

ABSTRACT

At the current time, interfaces between humans and machines use only a limited subset of senses that humans are capable of. The interaction among humans and computers can become much more intuitive and effective if we are able to use more senses, and create other modes of communicating between them. New machine learning technologies can make this type of interaction become a reality. In this paper, we present a framework for a holistic communication between humans and machines that uses all of the senses, and discuss how a subset of this capability can allow machines to talk to humans to indicate their health for various tasks such as predictive maintenance.

Keywords: human computer interfaces, predictive maintenance, acoustic analysis.

1. INTRODUCTION

A natural interface which can simplify the process of communication between humans and the next generation of machines is crucial to build systems that can deliver the best of their capabilities to easily solve a critical problem. However, current modalities by which machines interact with humans are very limited. When humans interact with machines, they use the tactile sense, e.g. use keyboards, either virtual or real, or swipes to provide their commands to the machine. The information provided back from the machines to the human user tends to be based on the sense of vision, in essence creating a display of pixels to communicate information to the users. While these modes of interactions are very useful, it would be possible to explore a wider range of communication options if we expand the scope of human computer communications to all the senses that humans posess.

We begin this paper with a brief discussion of the different senses of humans, and discuss how machine to human and human to machine communications can be enriched by using all five human senses. We then propose a layered approach for human machine communication, copying the ideas and philosophy of the layered approach used in computer communications. This is followed by considering one particular slice of the interaction, namely having the machines communicate their needs to humans using the sense of hearing.

2. HUMAN SENSES AND HCI INTERFACES

The interaction between humans and machines can be divided into two main types of interactions, one in which information flows from the humans to the machines, and the other in which information flow from machines to humans. The flow of information does not exclude interactive dialogue, because some of the information may be provided as an interactive exchange of different information flows. However, at the highest level of the semantics of the exchange, new information will either flow from the human to the machine, or from the machine to the human.

At this level of high level exchange, humans usually would instruct machines to perform some task, in essence providing them the task they need to perform. In addition to the commands that are being provided for a specific task, the human may also provide other high level guidance to the machine, e.g. define some policies [1] that they need to always conform with, or provide some required configurations. A policy may restrict the machine from interacting with only other types of machines, e.g. only with machines that belong to the same organization. A configuration is a more detailed set of information, and may include values for some parameters (e.g. how loud the volume should be). Depending on the machine, policies may be translated into configurations, and when the command for any specific task would come, the machine would conform within the restrictions put forward by the policies.

Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR VIII
edited by Tien Pham, Michael A. Kolodny, Proc. of SPIE Vol. 10190, 101900K · © 2017
SPIE · CCC code: 0277-786X/17/$18 · doi: 10.1117/12.2279891

On the converse side, at this level of high level exchange, machines report back to humans details of the outcome of commands or missions that they conduct. A machine may report back on the results, or try to draw the attention of the human to some situation that needs their involvement, or provide some statistics or reports on its conditions.

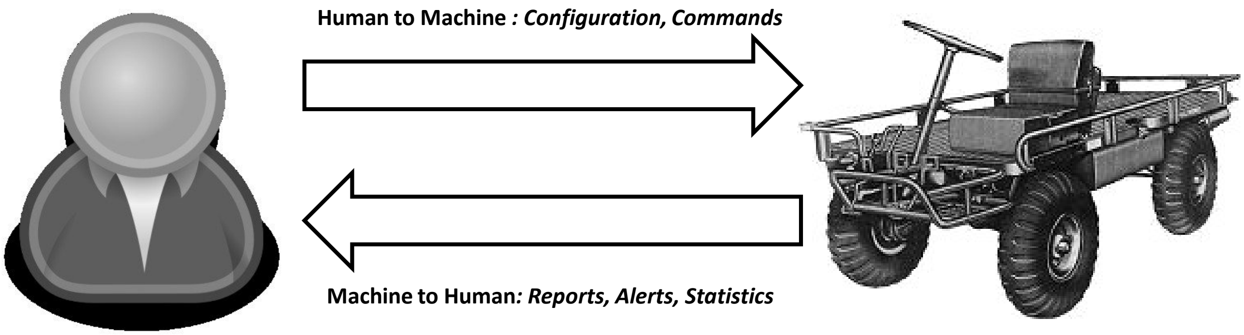

Human to Machine *: Configuration, Commands*
Machine to Human*: Reports, Alerts, Statistics*

Figure 1. Different modes of communication between humans and machines

The main focus of human computer interaction is to determine the best mode in which these two types of communications between humans and machines ought to happen. Towards this goal, we can note that the most prevalent mode for machines to provide information to humans is through the use of pixels, i.e. provide some display which the humans can see and understand the alert, report or statistics that the machine is sending. On the other side, humans usually would use some type of keyboard or touch-screen to provide their input (command or configuration) to the machine.

If we examine these two dominant mode of human machine interaction, we will observe that each of these modes is using only one of the five common senses that humans possess [2]. While some argument may be made for the case that the least number of modes be used for such interaction, and the current situation is just fine, a much more complex and richer set of human machine interactions can be provided if we can expand the interaction to more than just one sense that humans possess.

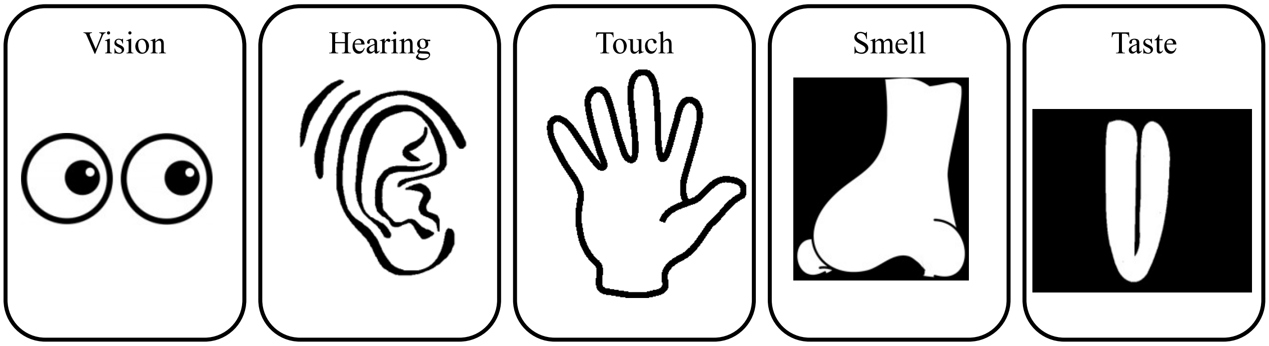


Figure 2. The five senses of human use

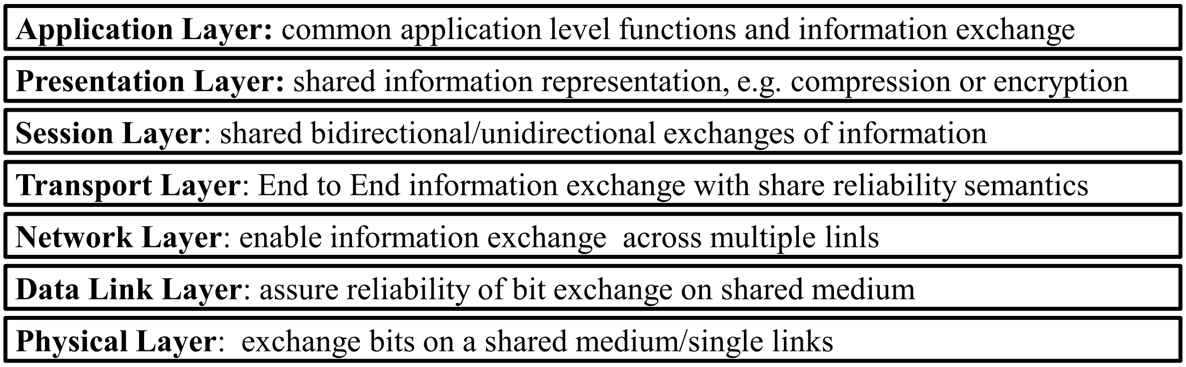
Most humans possess five senses shown in Figure 2 and they use these senses to send and/or receive information to other humans. These five senses, vision, hearing, touch, smell and taste, can be used concurrently for human machine interaction, just like they are used concurrently for a rich human to human interaction. For any human to human communication, a combination of one or more of the five senses are used to improve the exchange of information. In this paper, we examine how a similar multi-modal communication can enrich the interaction between humans and machines.

3. LAYERED ARCHITECTURE FOR HCI

To create a holistic multimodal communication between humans and machines, we borrow the best practices and design principles from a successful system for machine to machine communication. The machine to machine communication model that we use for inspiration is the 7-layer architecture for computer communications defined by the International Organization for Standardization. While the 7-layer OSI model has been mostly an academic reference, it serves as a good inspiration model to draw practical communication architectures which use the same design principles, but would typically use less than 7 layers.

3.1 Computer Communications Layered Architecture

The original layered computer to computer communication architecture proposed by OSI[3] is shown in Figure 3. It consists of seven layers, each with a distinct goal and objective.



Application Layer: common application level functions and information exchange

Presentation Layer: shared information representation, e.g. compression or encryption

Session Layer: shared bidirectional/unidirectional exchanges of information

Transport Layer: End to End information exchange with share reliability semantics

Network Layer: enable information exchange across multiple linls

Data Link Layer: assure reliability of bit exchange on shared medium

Physical Layer: exchange bits on a shared medium/single links

Figure 3. Computer Communications Layered Architecture

At each level, two communicating machines exchange information as protocol data units using a common protocol. Each layer insulates the upper layer from the details of the lower one, enabling situations such as allowing two machines with compatible transport layers to talk to each other even if they are using different network layer protocols.

The physical layer sends and receives raw bits over a shared physical medium. The data link layer ensures reliability of such transfers over the physical medium. The network layer removes the need for devices to be connected directly, enabling information to be router across multiple machines. The transport layer and in-sequence provides reliable exchange. The session layer provides multiple sessions back and forth as a one-way or two-way machine to machine dialogue. The presentation layer manages how information ought to be encoded, compressed or encrypted. The application layer provides a protocol to support the eventual purpose of machine to machine communication, such as file transfer or a database query lookup.

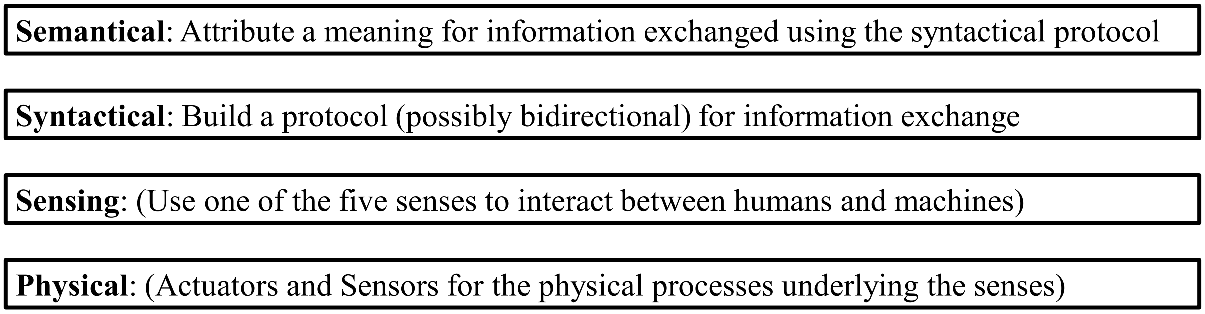
The key attributes in the layered design is the insulation among layers. This allows a transport layer protocol to run across multiple network layer protocols. The separation can also enable the a transport layer protocol to use more than one network layer protocols concurrently, e.g. transfer information over a wireless and a wired network concurrently. Such features can be used advantageously in human to machine communication as well.

3.2 Human Machine Communication Layered Architecture

Inspired by the layered architecture for machine to machine communication, we propose a similar layered architecture for a holistic human to machine communication which can span all human senses, which can manifest themselves in different manners. This layered architecture is shown in Figure 4.

The bottom-most layer in this human machine communication layer in the physical layer. The physical layer (which can itself encompass all the layers of the application stack as described in the previous subsection) consists of the actual mechanism used for human machine communication. This may include a network application protocol, or other ways of transmission of information. As an example, sounds may be transferred using some application level scheme from a device held by the user to a processor controlling the machine. The exact scheme depends on how the communication is realized. In some cases, the physical layer may consist of transmitting information using human-perceived sound, while in other cases the same information may be transmitted using ultrasound waves or via some special wireless interface. The task of the physical layer is to make sure that the communication of the upper layer information has been relayed over to the other side.

The next layer is the sensing layer. In this layer, the modality by which the human interacts with the machine is determined. The interaction modality may used any one (or more than one) of the senses available to humans. The concept of using more than one sense is consistent with recent observations in cognitive science area on sensory substitution capabilities of humans[4].



Semantical: Attribute a meaning for information exchanged using the syntactical protocol

Syntactical: Build a protocol (possibly bidirectional) for information exchange

Sensing: (Use one of the five senses to interact between humans and machines)

Physical: (Actuators and Sensors for the physical processes underlying the senses)

Figure 4. A Layered Architecture for Human Computer Communications

The syntactical level puts another mechanism overlaid on top of the senses. It defines a structure for interaction between humans and machines. The syntax defines schemes such as the interaction may happen using a sequence of Square and Oval images, or via other display, sounds or other combinations of sensory inputs. The top most layer applies meaning to the syntax.

This layered architecture provides a holistic way for humans and machines to communicate with each other.

4. EXAMPLE OF HOLISTIC HUMAN MACHINE COMMUNICATION

In this section, we discuss how the layered architecture described above can be used to create a new communication interface between humans and machines. In this context, we consider the typical interactions that needs to happen between the humans and machines in each direction, information flowing from the humans to the machines, as well as information flowing from the machines to humans. The information that is being sent from humans to machines includes the configuration and policies controlling the operation of the system, while the information that is sent from the machine to humans includes the current status, statistics and reports of the status of the machine (or collection of machines).

For the sake of illustration, we assume that the there is a single common semantical layer, in which the human asks the machine to operate in one of three modes (Normal mode, Emergency mode, or Exception mode), and the machine informs the human of its status (Normal Status, Degraded Status, Critical Status). In different situations, more complex semantical layers with a more complex mode of operation will be used. However, this simple model will serve well to illustrate our holistic communication architecture.

The semantical layer is mapped to a syntactical description. The syntax description is a representation of the modes at the syntactical layer. The may be defined in a structured language (e.g. XML), or in this simple case just be an encoding of the different modes in a binary format.

These sensing layer selects which of the human senses are to be used to communicate information back and forth. One or more modalities may be used concurrently for the exchange. If the speech mode is to be used, then humans talk and put the machine in one of the three modes based on selected phrases they utter. When using vision, a human may make a gesture in front of the machine to put in in a given state. The assumption here is that the machine would have a camera that can recognize the gestures and translate it into the desired mode. Similarly, in a tactile mode, the human may press the machine to provide its commands. In using the sense of smell or taste, the human may release specific scents in the air, e.g. using some chemicals, and receptors in the machine can detect those chemicals and put the machine in the right mode.

On the reverse side, when the human has to detect the mode of the machine, the human may listen to the sounds made by the machine to interpret its state (sense of vision), or look at some displays on the machine to understand its state (sense of hearing), or get some input in the tactile mode to understand its state. A tactile input may be obtained by the machine vibrating in different ways so that touching the machine provides information about its state. Similarly, the smells of the machine may indicate its mode, or the machine may exude some chemicals whose taste (sour, bitter, sweet) may indicate the state is in when tested.

Not all the modalities will be used for human machine information exchange, and taste/smell are not as commonly used for human machine communication as the other senses are. However, there is no reason to exclude those senses from the future generation of machine communication.

Once the sense modality is selected, each sense modality can be translated into specific function, such as the transmission of sounds using a specific sound frequency range (audile sounds from machines to humans, while ultrasound can be used from human to machine), or electromagnetic frequency. The specific way in which a sense is translated to machines and humans can vary and be highly dependent on the context of communication.

5. AUDIO BASED COMMUNICATION FROM MACHINES TO HUMANS

Having looked at the overall holistic model, we now consider a narrow slice of the human computer information exchange. We only look at the communication from machines to humans, and how machines can talk to humans intelligently. One part of this would be that the machine talks and informs the human in natural speech of its stage. While that is part of the scope of communication, we want to consider the larger part where the sounds being emitted from the machines during normal operations can be understood by humans to indicate machine state. These sounds could be in the audible hearing range, or be outside this range in the ultra-sound frequencies or infra-sound range.

For this narrow slice, we want to translate the sounds coming from the machine into one or more modes of the machine. The steps requires to do this translation is shown in Figure 5, and consists of three stages, the collection of sounds, the interpretation of sounds to ascribe meaning to them, and then displaying the meaning to the human. In this process, the hard part is the interpretation of existing sounds, which requires using machine learning techniques or other approaches to understand what the sounds in an environment mean. The collection of sounds is mechanical, and displaying the sounds to humans can be done in many ways, by sending a message, an audible beep of different types, or even have the machine use synthesized speech to talk to the human.

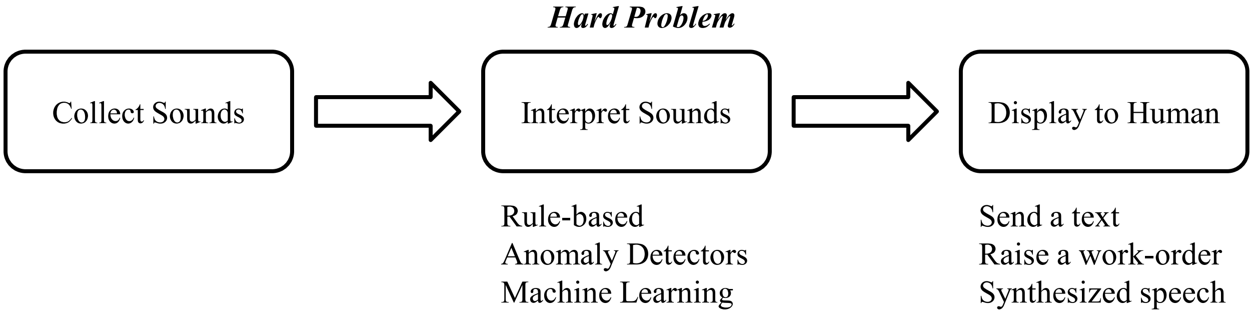


Figure 5. Audio Analytics for Predictive Maintenance

We will consider the specific domain of using audio analysis to diagnose problems with machines in the engine room of a building, although the same solution can be extended to many other machine scenarios. The solution consists of a listening device in the engine room, e.g. a sound collection application on a mobile phone, which is located in the engine room. The application listens to sounds and sends them over to an audio analysis system. The audio analysis system uses part of its listening period to determine what is the baseline of the system, and declare those sounds as normal. The sounds that do not fit into the baseline period as normal are characterized as abnormal.

The system for listening can be put into location without any prior data based training, and learns the ambient sounds in the environment on its own. As it detects the initial abnormal sounds, it marks the abnormalities by opening a ticket in the trouble ticket system, such as IBM Maximo. The trouble ticket system would then cause a human being to look into the situation of the machine, and take corrective actions. These corrective actions are logged into the trouble ticket system.

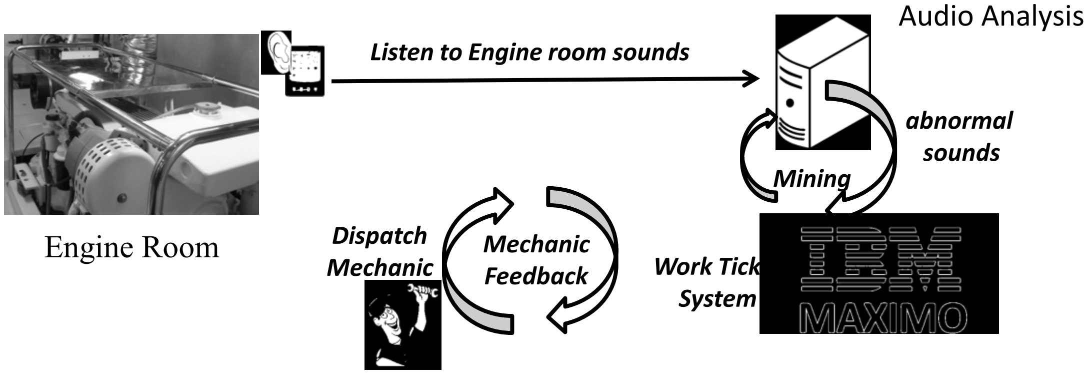


Figure 5. Physical layout of Audio Analytics for Predictive Maintenance

The audio analysis system mines the information in the ticket system at regular interval to see how the abnormal sounds are handled by the mechanic. The mining of the information allows the association of specific sounds to the right categories, so that the system can eventually learn to discriminate among sounds that characterize different types of situations, e.g. frayed belts, broken gears, or loose bolts will all produce different sounds which are distinguished over time and create a sound catalog eventually.

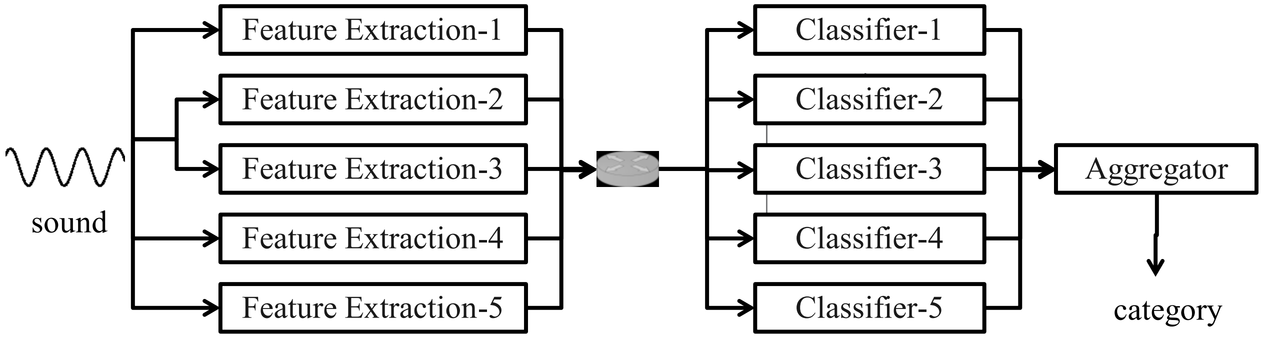


Figure 6. Ensemble Learning Model for Audio Analytics

Since each environment has its own dynamic set of sounds and are characterized by a specific environment, the system does not rely on any pre-existing classification or machine learning algorithm to perform its task. Instead, it uses an ensemble learning approach to get the best solution for any specific environment. During the base-lining phase, the system would extract multiple features from the sound system and also run multiple classifiers on each of the features. In each environment, the best performing classifiers and feature combinations are then used to make the appropriate classification mechanism which is customized and selected for that environment. This setup is shown in Figure 6.

Some operational aspects of the audio analysis system need to be considered to improve its usability. When an abnormal sound is observed, the anomalies will continue until the root cause is identified and the appropriate corrective action taken. Generating a trouble ticket on every anomalous sound will create too many trouble tickets, resulting in the state where the mechanics would turn off collecting and reporting of anomalous sounds. To prevent this from happening, the system needs to support policies to determine when anomalies are reported and when are trouble tickets created. The policies would only report an anomalous sound once it is observed for a minimum amount of time or a minimum number of times, which acts as a mechanism to filter out occasional sounds that may not be relevant to machine state. Furthermore, if an anomalous sound is reported in the trouble ticketing system, it does not open a new ticket on every new detection of the anomaly, but only one ticket per a policy-determined time-period, e.g. a ticket on a anomalous sound will only be created once a day. Furthermore, if the anomalous sound continues without any action, the policy would determine whether the existing trouble ticket is updated, or a new trouble ticket ought to be opened.

These policies are defined by two sets of people as shown in Figure 7. The domain experts provision policies which are applicable for all environments. The local building managers can define further policies which are specific to a building or environment where the solution is being used and deployed.

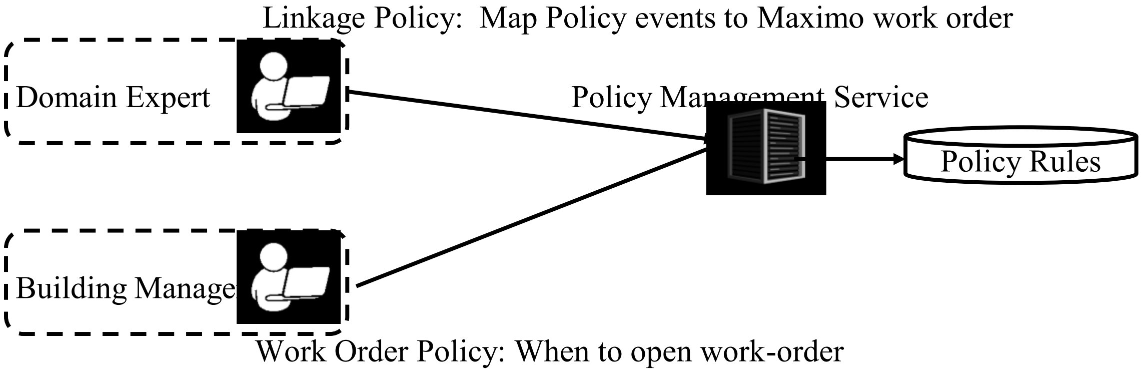

Linkage Policy: Map Policy events to Maximo work order
Domain Expert
Policy Management Service
Policy Rules
Building Manage
Work Order Policy: When to open work-order

Figure 7. Policy Definition Roles for audio analytics for predictive maintenance

Figure 8 shows how the policies defined in Figure 7 are implemented. These policies are used to provide the linkage between the trouble ticket system and the work order system. The work order system would only get new work orders as per policy defined by the building manager and the domain expert. On the other hand, information in work order is mined according to policies determining the frequency and complexity of such mining, and determine when the audio analysis service needs to be retrained with new labels and new data sources.

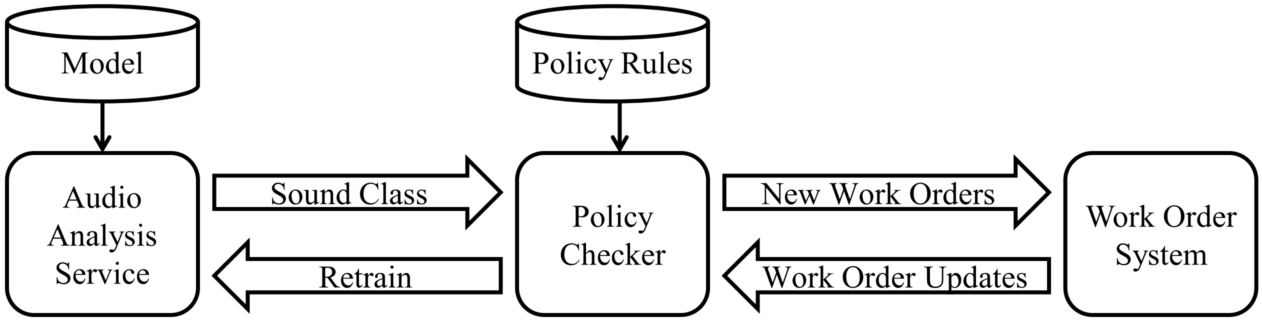


Figure 8. Run time Policy Enforcement for Audio Analytics

The solution provides but one slice (using one sense and one specific set of algorithms) of the holistic architecture to enable machines to use the human sense of sounds to talk to humans. However, this provides a proof-point that the approach works and can extended to support the complete set of humans senses to create a holistic human machine communication system.

6. ACKNOWLEDGEMENTS

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copy-right notation hereon.

REFERENCES

[1] Agrawal, D., Calo, S., Lee, K., Lobo, J. and Verma, D., [Policy technologies for self-managing systems]. Pearson Education, (2008).

[2] Rivlin, R., and Gravelle, [Deciphering the senses: The expanding world of human perception]. Simon & Schuster, (1984).

[3] Zimmermann, H., "OSI reference model--The ISO model of architecture for open systems interconnection," IEEE Transactions on communications, 28(4), 425-432, (1980).

[4] Bach-y-Rita, P., and Kercel, S. W., "Sensory substitution and the human–machine interface," Trends in cognitive sciences, 7(12), 541-546, (2003).