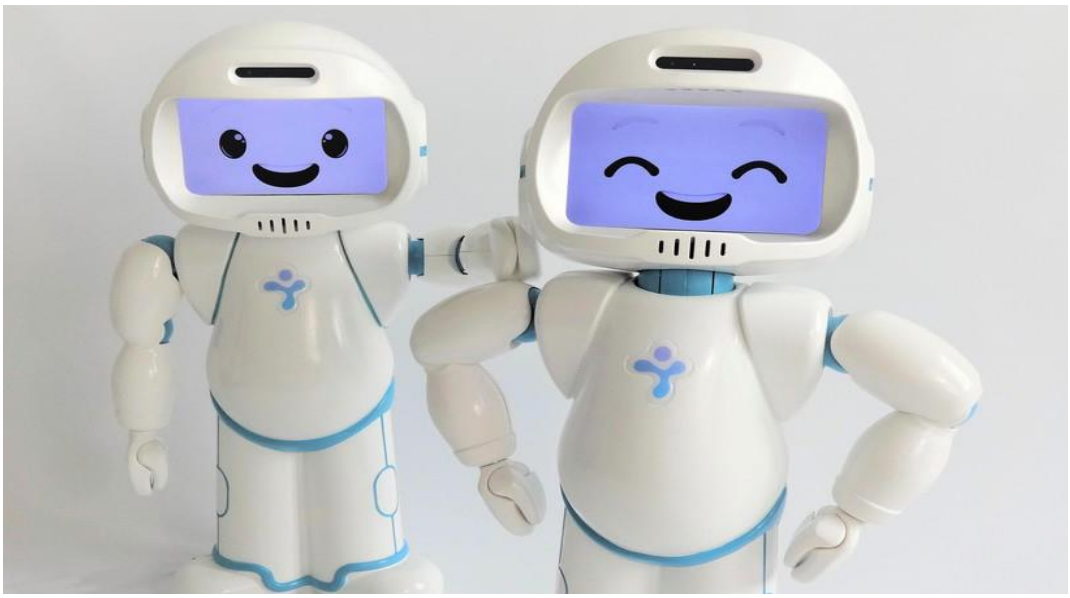
Designing sound for Robots

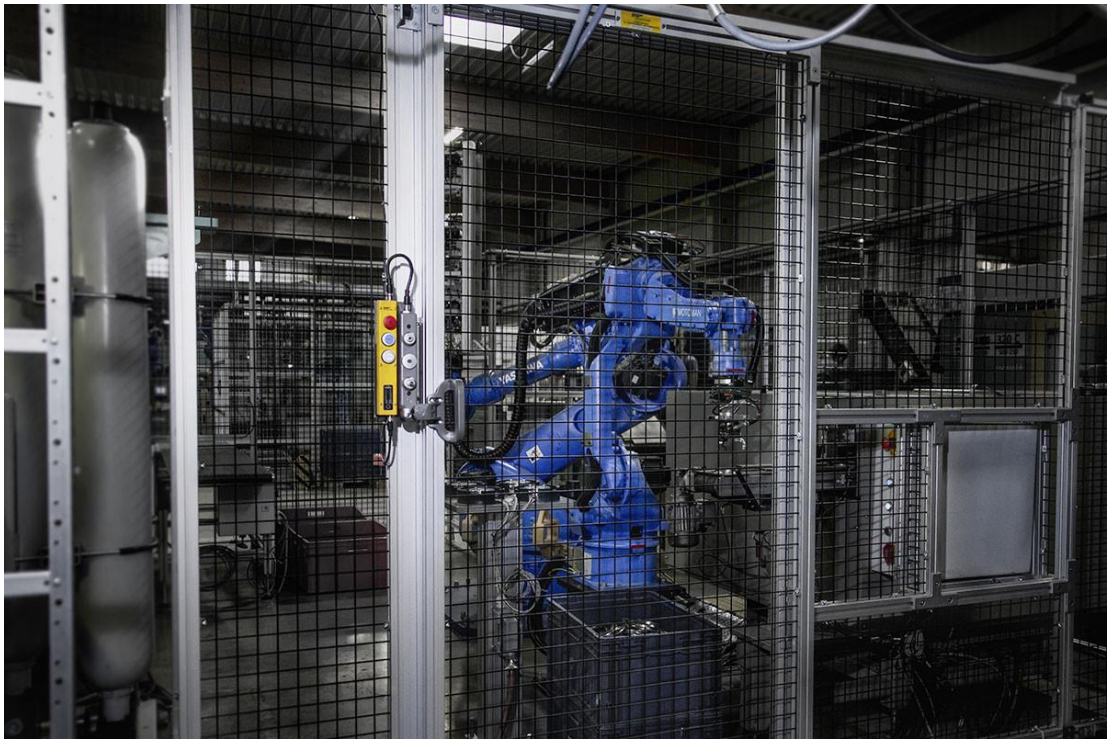
Iain McGregor
Online Sound Design MSc Programme Leader, Edinburgh Napier University



Robots can be thought of as servants, collaborators, professional best friends or even soul mates. The requirement for built-in sensors means that robots can potentially move between these roles according to the person they are interacting with. Whilst there are enthusiasts, reluctance to engage with robots is understandable when the differentiation between robotics and AI is often not fully understood, or even appreciated. Emotional engagement can be immediately affected by the auditory content of a robot. The wide range between a silent servant or a communicative companion can be reflected in the desired aural interaction. This process can be utilised in order to make a robot appear dumb or intuitive, according to the needs of the person interacting with it. At present when sound is designed for robots it is often speech centric or remarkably similar to that already utilised within digital devices and video games.



The frictionless interaction of speech input is becoming increasingly more popular in a wide variety of devices. Verbal communication has the advantage of conveying not only mood, but level of tiredness, character traits, and even some medical conditions. Medical practitioners already use speech as a diagnostic tool for Parkinson's, Heart problems and prostate cancer, which display as frequency notches in speech patterns due to voice deterioration, and these techniques have been built into commercially available software applications. The gender of a robot, and of virtual assistant technologies can be a major issue when a robot is perceived as subservient. Gender, or lack of one, can either be assigned or arrived at in order to make any humans as comfortable as possible in a robot's company. Gender neutral voices and action sounds can be tested in regular use, with the resultant responses used to guide an assignment by gradually trialling more male or female aspects until an optimal interaction is achieved. Alternatively, the robot could ask. "What would you like to call me?" The response can provide effective cues about which gender to assign, male and female names are generally obvious, with some exceptions. If the name chosen is more technology based, then a neutral gender voice can be applied.



There are a wide range of robots already working regularly in industry, and they are becoming ever more commonplace in medical and domestic environments. Robotics can take entirely new forms or be integrated into existing technologies in order to make them autonomous. In industry, robots are sometimes caged off so that no harm can come to any humans who share the workplace. In some instances, whole areas are completely out of bounds, such as docks or factory floors. The additional inclusion of proximity sensors wearable 'protective shields' can help remove or reduce this need for robot only environments, and the inclusion of appropriate sounds can further improve safety. Deliberately designing sounds with an approaching Doppler effect built in can help indicate when a robot or its appendage is coming towards a person, and might cause damage, making the interactions more natural as well as safer. If there is no danger, the device can be silent or make its normal operating sound. If the desire is to communicate that it is safe to be around the robot then a sound with the Doppler effect of a receding sound source can easily convey that there is no danger. An increase in pitch and volume is associated with a source moving towards a person, whereas a decrease in pitch and volume indicates a receding sound source. Individuals are regularly reminded of this convention whenever they are around audible moving objects and have learnt that those moving towards them usually needed to be attended to, whilst those receding can normally be safely ignored.



Everyday industrial items like floor polishers, which are often found in supermarkets and airports can be converted to become autonomous. The physical interfaces required for humans to operate them are often retained, so the impression that some have when first encountering these devices is that there is a malfunction and that it is operating out of control, minus its human operator. One option is to produce beeps, similar to a truck backing up, but this quickly becomes an annoyance for those who have to share the workplace environment. A second approach is to enhance the sounds that the device normally makes such as the circular brushes. That way they can be enhanced when the device is in close proximity to a person, so that if it slows down or stops entirely to give way for an individual, then it is clearly audible that the process has been interrupted. If the device is operating when no humans are present or at a proscribed distance, then the same audio reproduction technologies can be used to apply active noise cancellation, which whilst not being 100% efficient can still slightly reduce the overall levels.



Active noise cancellation can be helpful to reduce a robot's vibrations, which in turn can improve efficiency and accuracy of tasks. Surface mounted transducers may be utilised to vibrate out of phase in order to reduce any audible elements, making actions quieter. When it is not possible to cancel out a mechanical sound, piezo loudspeakers can be mounted internally to generate sounds that compliment and even extend the audible content. This can be used to provide confidence that a robot is capable of the task it is

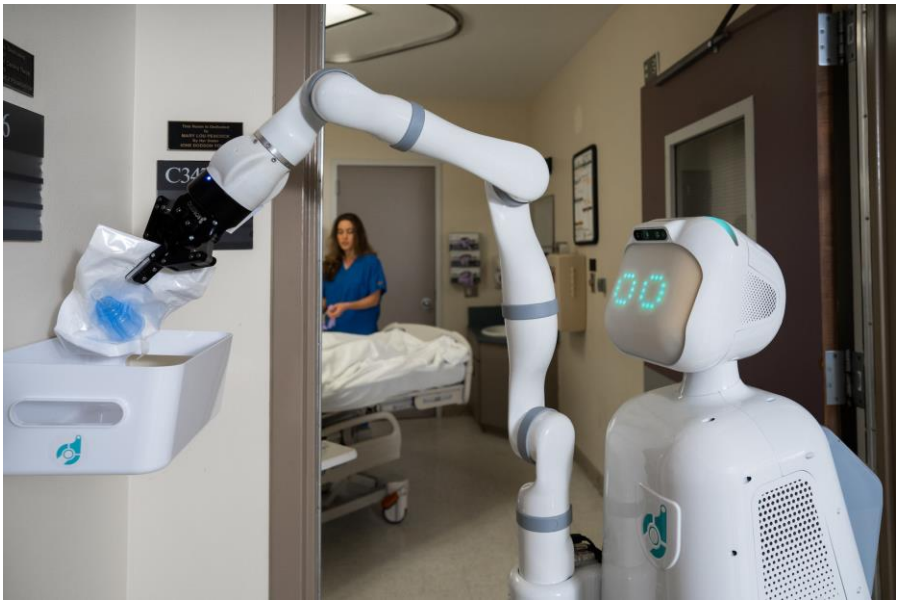
about to perform. So, when it is lifting someone it would make a different much stronger sound that when it is performing a delicate task that might hurt the individual being assisted. A ratcheting sound can imply that there is an inherent safety mechanism to prevent an action being accidentally reversed, or to prevent it from going too far. A smooth sound may provide confidence that no vibration will cause an item to be dropped. If the inherent nature of the sound producing robotic material does not inspire confidence, such as plastic, it can be replaced with a stronger more metallic sound to indicate strength. Similarly, if a level of elasticity is needed to convey delicacy then more yielding rubber type sounds can be used.



Customer facing robots such as those found in catering environments waiting on clients can generate sounds associated with heat to provide customers with confidence that the food is being kept hot. Conversely when a cold item is being conveyed, that it is being kept sufficiently chilled. These could be refrigeration sounds associated with cooling apparatus, slight high-pitched gentle fans or even quiet sharp cracking, if it is appropriate for the food item. Robotic chefs often have to use slightly different techniques for food preparation than human chefs and are sometimes operating in full view of the clientele. Heat, precision and cleanliness are all simple concepts to convey aurally. Sizzling has long been associated with heat, and latterly the Fajita effect, regular timings of any movements which convey the correct texture imply accuracy, and steam, suction or brushing sounds communicate a freshly prepared hygienic surface. For large scale industrial robotic lines then, similar principles can provide confidence that hidden aspects of a production line are operating correctly.



The uncanny valley is a much-cited issue associated with androids. The intended human appearance can evoke unwanted emotions, but with monitoring, sounds can be used to vary the level of perceived realism in order to make humans feel more comfortable when interacting with human like robots. Behaviours such as eye contact, proximity and pronoun use when addressing an android can be analysed in order to know whether to make the android appear aurally more or less realistic. This approach might sound counter intuitive but communicating the artificial nature can assist with engagement so that discomfort can be gradually overcome through repeated exposure. After an acceptance has been reached a more natural realistic set of sounds can be utilised, which partially obscure any artificiality. Random disharmonic artefacts with a wide dynamic range can be included when on the artificial end of the scale, counterbalanced by even-order harmonics and a more predicable dynamic range when naturalism is desired. This can be a gradual transition applied to speech, that can be further extended through the use of more or less formal language, as well as false starts, errors and repetition. These artefacts can also be used to hide or emphasise processing delays, in a similar manner to when humans are trying to think of an appropriate response. Fricatives, sibilance, lip smacks, breaths and glottal fry can all be included after the initial discomfort has been overcome. The underlying hardware can also be emphasised or deemphasised accordingly, so that pneumatics can made louder, or distracted from by overlaying vocal artefacts.



Medical robots range from those used in surgical procedures, through to medicine dispensing and rehabilitation, amongst others. Hospitals are already considered noisy spaces, due to the well-established practice of medical alarms, highly reflective surfaces, visitors and staff, many of whom make active use of mobile phones. The World Health Organisation daytime recommendation of 35 dBA for the benefit of patient recovery is typically exceeded by at least 25 dBA, as is the 30 dBA night time level. Critical patients have been shown to experience aural disturbance every six minutes during the night in some hospitals. Robotics in a medical environment can dramatically assist in the lowering of levels through a variety of techniques. The first is through active monitoring, most of the existing artificial alert sounds can be pitch shifted up into the ultrasonic range so that they can then be translated back down to an audible frequency in the appropriate location, such as a nurses' station or when the desired listener is within range. Many robots can also move themselves to locations where they can either provide masking sounds or amplify an essential sound in order to provide an auditory bread trail to decrease staff response times. When essential sounds still have to be heard on a ward complementary sounds can be generated by a robot, or similar device so as to make any alerts less stressful for those who are not the intended auditors. This can be made more fun for paediatric wards using wildlife or fantasy sounds, or more naturalistic for adult wards, with the further benefit of providing privacy when needed, like an acoustic curtain.



Robotic carts/trolleys are not confined to hospitals, but they are increasingly common. At present the auditory interactions are mostly through speech, but these can be altered according to the urgency and sensitivity of the items being transported. When a clear path is essential attention can be drawn through creating an artificial motor rotation, near silence can be adopted for less time-sensitive journeys. Auditory warnings associated with the cargo can be transmitted as it approaches the intended recipient so that preparation can be made in advance, without having to actively look out for its arrival. When moving through paediatric wards, the cart can become a form of auditory spectacle, from an imaginary unicorn or even a spaceship. The inherent nature of the robotic carts is that they tend to move more smoothly than manual trollies so in themselves they already reduce the level of noise within a medical environment.



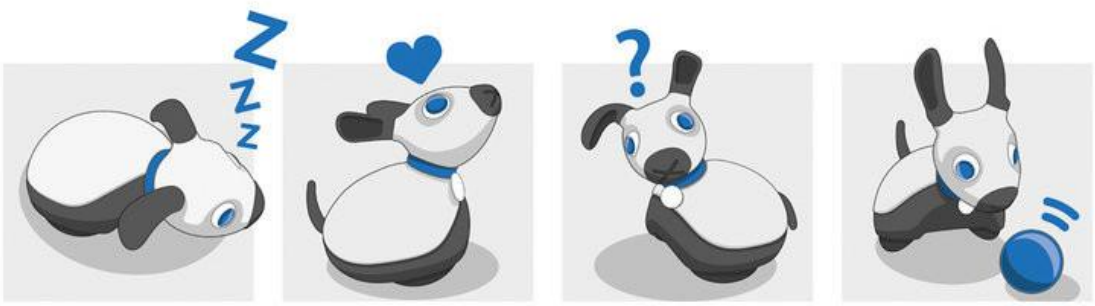
Companion robots were initially designed for hospital environments and have been used to comfort children and the elderly alike. They provide distraction and can even promote the reduction of cortisol and increase of oxytocin, while remaining clinically hygienic. Soothing music has been repeatedly shown to help patients reduce stress in hospitals, which has often been provided by hospital radio services. These benefits are not confined to music, the sounds of cats purring, gentle breathing, waves and gentle wind all have shown similar effects. An optimal approach to the design of companion robots is to maximise the level of interactivity so that a high level of engagement is achieved. Sound choices can be cycled through until the optimum human response is achieved. The animal kingdom is often used as inspiration for personal care robots and those designed to engage children, in order to increase the level of bonding, and perceived personality. However, little is generally done to auditorily mask the mechanical nature or to convey the underlying processes, which can provide much needed confidence for those interacting with robots.



Dementia active monitoring can be applied using companion robots so that the auditory interventions can be adapted for maximum effect. Simple tasks such as reminding people to move, sit down or eat, can either be achieved through speech synthesis, or by associative sounds, such as a microwave 'ding' or food preparation sounds. These can be more traditional sounds associated with their own youth, so that the level of recognition is improved. These can also be extended, so that tasks such as cooking or even changing a lightbulb can be assisted by a robotic companion, providing a verbal commentary and supportive sounds when needed.



Companion robots are also available as toys and are becoming remarkably similar to those designed for medical use. Science fiction and nature are both popular tropes for sonic design, talking dolls have been available since the 19th century, and are still readily accepted by the young. Decades of science fiction films with robots has created an expectation for robots to sound futuristic, which some definitely do, although often with a cute, friendly edge. The microphone arrays that robots use allow an auditory contextual understanding. One technique adopted from smart speakers is the companion app that listens to someone reading a known book and then automatically plays back sound effects or music to punctuate the story. But this can extend to any auditory interaction, so if a child laughs the robot can join in, or if crying is heard then soothing sounds can be generated. They also have the advantage, if having been suitably programmed, of being able to transmit the sound to any concerned adults, through whatever device they have access to.



Preprogramed personality traits are popular, and even more so when they have the ability to develop. Stroking can produce purring, being ignored, or mistreated may lead to negative verbalisations. Literal interpretations are often a starting point. Novel associative sounds that are reminiscent often work best, as they do not draw attention to the artificial nature of the source, so that it is not perceived as being less than a 'living creature'. Analysis of local languages for sounds can be an intuitive method of designing sounds. Spearcons, which are dramatically speeded up spoken words, have proved effective in interfaces for those not able to visually see a screen, but they can also be used to provide spectral, dynamic and temporal cues that can be applied to sound effects. The emotional content associated with vocal delivery is commonly perceived and can even be used as an input to provide auditory mirroring, so that the sounds generated by a companion robot are even more closely aligned with the person it is interacting with.



Many auditory cues are accurately interpreted cross-species, crying is the most cited, but there are others, the only real constraint is ensuring that the spectral content falls within the auditory range of the listener. Social media accounts can be utilised beyond advertising in order to provide candidate animals for mimicry. A strong interest in for example penguins or cats can form the basis of a customised auditory experience, especially relying on previous exposure to create meaningful cross technology communication. Phobias can also be avoided, and even when information is not available then grouping of similar profiles can fill in some of the gaps. Tropes from other preferred media such as video games or films can also be applied to extend the auditory palette. The auditory feedback loop can then be applied to alter sounds according to their reception, in order to maximise the level of end user comfort.



Some principles of sound design have already been trialled within robotics. Musical sequences are often included in the auditory output of robots, but the concepts apply to all sounds. Rising pitch of non-verbal utterances are considered argumentative or angry, and slow decreasing pitch as a form of hesitation or sadness. These are both representative of neutral valences, just with positive or negative levels of arousal. Neutral pitch is perceived as difficult to understand in terms of emotional content. High, variable pitch and dynamics sounds are perceived as excited. Duration and speed of a sound is, unsurprisingly, related to being energetic or lazy. Happiness has a richer balance of upper harmonics, sadness has more emphasis on simpler lower frequencies, with little variation. "Beeps and chirps" are one way of describing the resultant sounds, which are popular in both films and toys. Four main approaches have been adopted to date within robotics: Gibberish speech, musical utterances, non-linguistic utterances and paralinguistic utterances. All of which are effective to varying degrees, but almost all become more successful through repeated exposure.



It is also important to know when a robot is actively monitoring and ready to perform, or when we really need to know that it has powered down and there is absolute privacy. There are so many differences dependent on context, there are also all of the upcoming autonomous items such as vehicles, but many of the smart technologies will have some level autonomy, and we can use sound to reassure users that they still have control as well as to minimise the perception of risk. Truncated sound helps communicate that a device is inactive and will require a physical action for it to start up again. Standby would use a similar sound with a natural slow decay, with a corresponding gentle rise for a return to full functioning status.



Rather than slavishly duplicating human, animal or machine sounds, robotics offers the opportunity to develop an entirely new form of auditory communication. Extremes that were not appropriate in other devices or creatures can be highly advantageous for immediate engagement, to warn of potential damage as well as encourage close personal contact. The interactive loop can be used to audition sounds to gauge their reaction and adapt accordingly. The ability to connect via the internet to a larger database or AI ensures that an expanding range of sounds can be explored, as well as to allow an element of mimicry, either of the humans or other animate or inanimate sound sources within an environment. Source identification does not have to be confined to purely visual elements, the level of engagement with other identifiable objects can also be captured in order to refine a robot's communication.



A robot can be a highly intuitive friend, who always wants to make everyone at ease, in as an efficient and effective aural manner as possible. The robot listens and provides an auditory backdrop to convey both its intentions and its own level of engagement. If the requirement is for an efficient assistant then silence, with only essential simple sounds can be expressed. If an engaging companion is preferred, then a fully adaptive and intuitive auditory interaction can be shared. A well designed robot has the potential to have a frictionless natural auditory interaction with those around it, so that irrespective of the context a harmonious existence can be achieved.

© Iain McGregor All Rights Reserved Find more at www.soundformovingpictures.com