6

Consequential Sounds and Their Effect on Human Robot Interaction

Aimee Allen, Richard Savery and Nicole Robinson

6.1	Introduction		101
6.2	What are Consequential Sounds?		102
6.3	Psychoacoustics: Human Perception of Sound		103
	6.3.1	Human Perception of Consequential Sounds	105
6.4	Product Design for Consequential Sounds		106
6.5	Consequential Sound Spectrums of Robots		107
	6.5.1	Existing HRI Research Involving Consequential Sounds	109
	6.5.2	Consequential Sounds of Different Robots	110
	6.5.3	Case Study: Comparison of Consequential Sounds across Different Robots	112
6.6	Capturing and Altering Consequential Sounds		115
6.7	Design Implications for Consequential Sounds in HRI		120
6.8	Research Potential for Improving Consequential Sounds		121
6.9	Conclusion		123
	Attributions		123
	Bibliography		123

6.1 Introduction

Robots are more commonly being deployed into human-occupied environments such as workplaces, public spaces and homes, leading to humans and robots working together in close proximity. Effective human–robot interaction (HRI) is a key component in the successful use of robots in human spaces, and careful interaction design and deployment helps to create frequent or long-term interactions that are beneficial to people. Interaction design spans across both visual appearance and audible sounds, with most existing research focusing

DOI: 10.1201/9781003320470-6

on design for visual components, such as the robot physical appearance, communicative gestures or facial expressions.

Sound is a highly influential component that underpins successful human–robot interaction, given that the human brain often prioritises sound inputs over visual inputs [16]. Sound in human–robot interaction can have strong benefits, such as helping people establish a sense of proxemic comfort and localization of other agents within our environment [5,38]. Sound can also have negative consequences, such as if a sound is too loud for the situation, or the wrong sound is present during the interaction, leading to a sense of confusion or annoyance [24, 31, 41]. Despite the increased focus on sound research in robotics, the study of most non-natural-language robot sounds, including consequential sounds produced by the robot's actuators, is comparably rare within HRI research. This is despite the noted importance of non-language sound in human-human interactions [45] and the fact that consequential sounds are prevalent in almost every robot and thus every human–robot interaction.

Consequential sounds are of concern in human robot interaction because they are often perceived negatively by most people [10, 11, 24, 36]. As consequential sounds are currently extremely under-researched as noted in several publications [5, 9, 10, 23, 33], there is an important research gap around how effective sound design for consequential sounds will contribute to robot acceptance by people including their willingness to purchase, use, and work long-term with robots.

This chapter explains robot consequential sounds, and the resultant impact on human robot interaction. Different consequential sounds across a variety of robotic platforms are demonstrated, and the effect robot consequential sounds have on HRI are discussed, including techniques and implications when designing for consequential sounds in research or production soundscapes to improve human–robot interaction success.

6.2 What are Consequential Sounds?

The term 'consequential sounds' was first coined for the purposes of product design as "sounds that are generated by the operating of the product itself" [18]. In other words, consequential sounds are the **unintentional noises** that a **machine produces itself** as it **moves and operates**. Consequential sounds exist for any machine that has moving parts as they are the audible sounds that are generated when different mechanical parts or actuators operate together to make the machine work. Due to the machine's specific functional design, the machine can not perform normal operations without making these sounds.

Consequential sounds originate from **vibrations of components**, generating forces which create a sound compression wave. These vibrations can be caused by the friction of turning parts, including friction within fast-spinning components known as ego-motion sounds [42], interactions with air/fluids in pneumatic/hydraulic actuators, resonance in other materials (e.g. chassis or joints) or other vibrations of actuators. Some examples are the whirring of a refrigerator condenser, soft static ticking of an idle television or computer screen, the spinning thumps of a washing machine's large motor or the buzz of a vibrating electric shaver or toothbrush [26]. Consequential sounds are generated internally by the machine, rather than by what the machine interacts with, and most often will be audible to nearby people. These sounds exist before any additional extra sounds are implemented, such as the use of audible tones to signal to the user that the machine has completed a task.

Many consequential sounds consist of two main components; broadband noise with some specific narrow-band harmonic noise. Repeatedly spinning components (such as motors) tend to generate broad spectrum sounds, with narrow tones/harmonics often coming from fast bursts and sudden stops of actuator motions. The precise spectra of these have been shown to vary not just by actuator type but also quite dynamically with changes in the rotation speed of the motor [43].

There are two key criteria to help distinguish consequential sounds from other machine generated sounds:

Criteria 1: Consequential sounds must be **unintentional**. Therefore, this criteria will exclude all types of intentional sounds that a machine may produce. For example, sounds that are intentionally programmed into the machine to interact with a human, such as sounds that produce verbal speech, non-speech sound, music, or beeps/tones, including sounds used to communicate a state, function or affect of a robot. Any sound that an engineer or designer has chosen to add to a machine and use through a speaker are not consequential sounds.

Criteria 2: Consequential sounds must be **produced by the machine itself**, rather than something the machine is interacting with at the time, so this excludes noises such as vibrotactile sounds. Vibrotactile sounds are the noises generated from the vibrations which occur when two objects touch. For example, if the machine makes contact with the environment, or interacts with another object or person [21]. In the case of robots, this could include the sound produced when a robot walks or rolls across a surface, grasps items, or collides with a basket when picking an object. Vibrotactile sounds are not consequential sounds, however these sounds are likely to be perceived similarly to consequential sounds. Additionally, the same interactions which cause vibrotactile sounds often alter consequential sounds through changes in strain and friction inside the actuators.

6.3 Psychoacoustics: Human Perception of Sound

A human's perception of a robot's sound will be critical to the user experience and success of robots that interact with people. Thus, a basic understanding of how people perceive sound is crucial for anyone designing or researching

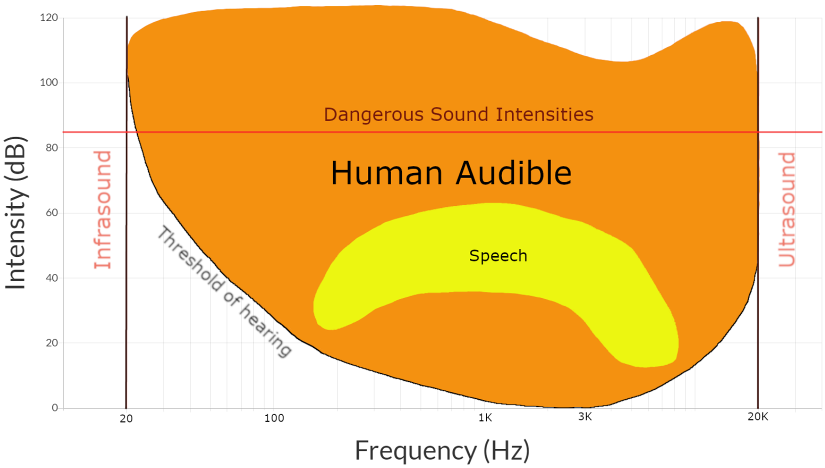

FIGURE 6.1
The intensity(dB) of sound which can be heard by humans (shaded region) across the typical human audible range (20Hz–20kHz). Human speech is typically centered around the middle of this range (banana-shaped region).

robots intended for human robot interaction. Psychoacoustics is the study of human sound perception and audiology, specifically the human psychological response to different sounds. The normal human hearing range is between 20Hz and 20kHz [5, 7], with this frequency range decreasing as a person ages. Humans find it easier to hear sounds in the middle of this spectrum (around 2kHz), which is the frequency range of human speech. Toward the outside of the human audible range, a higher volume (intensity) of sound is required to be above the threshold of hearing and thus be heard [5] (see Figure 6.1). Hertz (Hz) have an exponential relationship with the human perception of pitch, with each doubling of hertz (such as 440 to 880) representing an increase of one octave. Sounds of any frequency above a certain intensity (roughly 85dB) are dangerous as they can very quickly damage hearing.

Sound contains two main **objective components**: frequency (how fast the sound is oscillating in Hz) and intensity (strength of power in dB). Sound contains three main **perceptual components**; pitch, loudness and timbre. Pitch is how high or low the frequency is perceived, loudness is how strongly the intensity is perceived and timbre is a mixture of most of the subjective elements of sound such as how abrupt (sharp) or how pleasant the sound is to listen to [8]. Sharpness of a sound tends to decrease with a wider bandwidth of mixed frequencies. In sound design, it is critical to measure and understand the objective physical sound elements (frequency and intensity), and to also gauge the subjective sound elements as to how sounds are heard and perceived by humans (pitch, loudness, pleasantness) [12].

In general, humans find broadband noises (less sharp sounds) more pleasant than pure tones [8] and particularly favor those sounds where the frequencies (harmonics) are evenly spaced across the human-audible range [5]. Consequential sounds made by continuously spinning actuators (such as motors driving wheels of a mobile robot) tend to be mostly broad spectrum. People often respond more positively to rhythmic predictable sounds than sudden acute sounds such as alarms, car horns, or the sudden startup noises of a machine [16]. These sounds tend to disrupt focus by eliciting an instinctive danger or warning response. Sounds of this type when produced by robot actuators have the potential to both distract co-inhabitants, or ruin co-ordination between human–robot teams [16] by risking interrupting the natural human psychology to form spontaneous synchronised movements and emotional connections with a coworker.

People often encounter unpleasant sounds, with noise pollution being very common in many human environments [41]. Often the noise pollution is at least partially due to consequential sounds such as noise from cars, aeroplanes and trains; and machine consequential sounds such as air-conditioners. These persistent, annoying sounds can have strong mental and physiological effects on people including; annoyance and irritability (with social consequences), high stress (increased cortisol levels), anxiety, cognitive impairment, distraction and reduced productivity, and sleep disruption [4, 14, 41]. If robots become as well utilised as these other technologies, then their prevalence, and thus regularity of their consequential sounds, may lead to a noise pollution classification and negative perceptions. An example is the increasing noise complaints related to drones being used for delivery in residential areas [15].

6.3.1 Human Perception of Consequential Sounds

How consequential sounds are perceived, and thus how they could be used or altered, will vary across a number of individual and environmental factors. Common negative perceptions of consequential sounds are often not entirely due to the objective intensity levels or spectra of the sound, but rather the perception of the noise varies depending on the context of environment and tasks being undertaken [24, 31, 41]. Trying to work or rest at home typically leads to people perceiving sounds as louder and more annoying than when they are doing leisure activities [41]. For example, abrupt changes in sound (sharp sounds) can contribute to causing distraction or confusion, so abrupt consequential sounds can lead to a negative response in these contexts [38]. Type of environment, time of day, use of space, and the interactions between agents (both human and robot) within that space are all known to change sound perception [24, 41]. Context may include variables such as current task or mindset of a human (concentration versus socialising), the environment they are within (work, public or at home), as well as the expectations of the robot and its purpose (what is the robot assumed to be doing) within the context. The same high intensity machine sound can have a negative effect (noise pollution) whilst concentrating or sleeping, but may be seen as positive whilst interacting with the machine, such as correlating with power in a car or motorised tool [24].

Another factor is personality driven preferences of an individual listener [12]. Some people are consistently more or less annoyed by sounds than others based on their global 'noise sensitivity' [31], meaning that some people are more sensitive to sounds and being interrupted by them. Personality attributes such as level of introversion/extraversion are known to contribute to an individual's global noise sensitivity. This means consequential sounds made by robots may be viewed differently by different individuals.

There is also a spatial component to consider, given that sound propagates through space and can be experienced differently by different occupants [24] (a further study on spatial sound can be read in Chapter 4). People tend to prefer sounds from embodied (particularly moving) agents to convey a sense of proxemics, i.e. a sound that denotes their presence and relative positioning [38]. Humans are familiar with sharing environments with biological agents, and are accustomed to regular, rhythmic sounds such as breathing or rustling sounds that other people or nearby animals make [16]. In addition, it is often best when auditory and visual systems reinforce each other, and as such it may be unnerving to see but not hear a robot's presence, or to hear its presence more strongly than it is visually apparent [16]. Altering consequential sounds to create a consistent proxemic sound may increase comfort levels and therefore be less likely to create violations of 'personal space'.

6.4 Product Design for Consequential Sounds

Consequential sounds have been studied as a part of product sound design for decades due to their influence on a person's opinion of a product, and thus a person's likelihood of purchasing or using the product [6,18,19]. To achieve the desired product sound dimensions for the intended user perception, researchers, designers and engineers often focus on individual components and design principles to improve the sound experience. Product design theory describes several dimensions of product sound that are relevant to consider: strength, annoyance, amenity, and information content [19]. Strength or magnitude includes both objective dB intensity ratings as well as a loudness perception. Annoyance is a perceptual element consisting of factors such as sharpness, roughness and noisiness. Amenity or how pleasing the sound is to a person is very subjective and thus a challenge to measure, but includes elements such as rhythmic/regular sounds, harmonious qualities, and contextual appropriateness of the sound. Information content refers to the properties of the sound which communicate what or where the product is and its current task, performance and condition, and often consists of many intentional sounds alongside the consequential sounds.

To reduce the impact of consequential sounds during sound design, researchers, designers and engineers often focus on sound dampening or cancelling methods within specific sub-components. Example sound reduction and cancelling methods include: passive noise reduction by adding sound adsorption layers to enclose and dampen actuators, and active noise control, which makes use of added small sensors and speakers to help counter-act machine-like sounds. For example, cars often make use of active noise control techniques to reduce a narrow band of undesirable low frequency noises such as ground sounds and engine hum [30]. Microphones and vibration sensors are placed in multiple locations around the car to measure noise signals, and the car radio speakers each play sound of the opposite phase to cancel the measured signals. Many very similar technologies exist which are capable of generating opposite phase signals 2ms after the undesirable road noises are detected. Another common noise reduction technique is carefully designing the materials or shape of components to reduce and alter sound to a more pleasant spectrum. For example, in order to combat increasing complaints of noise pollution of drones in residential areas, researchers investigated using odd numbers of blades on propellers to generate a more even broad spectrum of noise [17]. Additional techniques used for managing consequential sounds are expanded on in Section 6.6.

6.5 Consequential Sound Spectrums of Robots

Robots produce a large spectrum of consequential sounds. These sounds must be considered not only in the design of the robot but also when planning implementation of a robot into physical spaces, as well as subsequent perception of the robot by users during human–robot interactions. The type of robot and its specific design are large factors which influence the spectrum of consequential sounds generated by a robot. There are clear differences in the type and frequency of consequential sounds generated by different robot form factors and motions. For example, humanoid robots moving their limbs to communicate with people, slow spinning motors to locomote a wheeled robot, walking actuations of legged robot, fast spinning props on a drone, or industrial robots conducting pick and place activities. Other significant contributors to the types of consequential sounds generated include: the number and type of actuator (motors etc), actuator positioning within the robot, and the shape and material composition of components surrounding or touching the actuators.

A large challenge for improving consequential sounds produced by robots is simply the shear volume of different actuators and robots that need to be addressed. Furthermore, an individual robot's consequential soundscape will also change over time, depending on its current operation and composition. For instance, a robot moving at rapid speeds at infrequent times compared to slow constant motion. Some of the common variables that impact and

TABLE 6.1
Variables that impact consequential sound generation.

Variable	Description of Effects
Component choice (actuator type and specs)	Different actuators create different base sounds. In general, higher power creates louder sounds. Electric motors are often the quietest actuators for their size/power
Product layout (how components are arranged relative to each other)	Components can generate consequential sounds by passing compression waves through each other, thus component relationship matters
Turning speed of actuators	Typically faster speeds generate more even sounds. At high speeds, these sounds can be negatively perceived as a frequent buzz
Full revolutions (continuous spinning) versus smaller angle motions (e.g. precise positioning of stepper motors)	Continuous rotations generate smoother sounds. Stop-start motions of small angles generate acute sounds
Material choices for inactive components (especially housings/chassis)	Certain shapes and surfaces can cause additional vibrations, resonate or amplify the actuator noises
Current power levels (important for DC powered devices)	As the battery drains, the voltage supplied to other components changes, and this alters the consequential sounds produced. Variable power supplies create variable sounds. Straining actuators can become louder and less pleasant if power is insufficient
Temperature of environment and components	Temperature changes how much strain the actuators are under, as well as expanding or contracting certain elements causing a change in friction generated sounds
Surfaces standing on/or objects interacting with (excludes vibrotactile sounds created by contact with the objects themselves)	Properties of contact object changes the exertion of the actuators and thus the sounds they produce
Imperfections in the robot as it ages e.g. dents in chassis	Minor shape differences change the acoustic properties of any resonance and sound amplification of the other consequential sounds

alter consequential sounds can be found in Table 6.1. Some variables or components have a larger effect than others on the spectra of consequential sounds produced. However, even minor changes can have an impact on the interaction experience, given that humans are able to perceive these sound differences [32], which has been shown in existing research to influence their opinions of a robot [5, 9, 10, 23, 33].

6.5.1 Existing HRI Research Involving Consequential Sounds

One of the first robotics papers to focus on "consequential sound" [24], used an online Amazon MTurk study to compare non-contextualised DC servo motor sounds on their own without robots. Participants were asked to compare videos of pairs of DC motors with dubbed sound i.e. with the 3D room ambient sound qualities removed. Consistency was found within participant ratings for preferred sounds, but sound preferences across participants were not consistent, suggesting a need for identifying more globally accepted pleasant or neutral noises. A second experiment overlaid consequential sounds from low versus high quality robotic arms onto videos of a high quality KUKA desktop robotic arm to investigate differences in perception of the robot when consequential sounds changed [36]. All consequential sounds showed a reduction in aesthetic ratings compared to silent videos, and the consequential sounds from the higher quality robot (which matched the video footage) correlated with higher ratings for competence of the robot. Another video-based study attempted to break down how specific variance in sound attributes (intensity and frequency) of consequential sounds affects perception of robots [46]. Videos of a UR5 robot arm had their natural consequential sounds manipulated up and down in terms of sound intensity (volume) and frequency (pitch). Results suggest that quieter robots are less discomforting, and higher frequency sounds correlate with positive perceptions such as warmth.

Several in-person HRI experiments have shown that consequential sounds can interact with other sounds or robot gestures to confuse the interpretation of affect [9, 11]. In one study, the low-frequency consequential sounds of a NAO humanoid robot created a strong arousal, negative valence affect [11]. This led to the robot's other sounds or gestures being perceived as frustrated, regardless of the intended affect. Another study examined whether augmenting consequential sounds of a soft robotic manipulator could change how happy, angry or curious the robot seemed [9]. Participants wore headphones which both deadened existing DC motor and pneumatic consequential sounds, and played additional affective sounds to supplement the movements of the robot. These changes to the consequential sounds altered the perceived valence of the robot to be happier, less angry and more curious.

Other research has investigated improving consequential sounds of a micro-drone within a naturalistic indoor home-like setting [44]. Three sound conditions: addition of birdsong, addition of rain sounds and no additional sounds, were tested at three distances from participants: near, mid and far. The masked consequential sounds were preferred at all distances as they were perceived as more pleasant than the unaltered consequential sounds. Which masking sound was preferred varied depending on the participant's distance from the drone, as well as their prior experience with related sounds i.e. existing associations with birds and rain.

Some non-sound focused HRI studies are also beginning to note the interaction effects of consequential sounds on other elements. In one study [37], participants deliberately limited robot motions specifically to avoid generating disliked consequential sounds, citing reasons such as "It even comes down to just how loud the motors were the first time it moved, that's very abrupt in a sonic way". In another context, autonomous vehicle researchers found that people use familiar consequential sounds, such as car engine noises, to locate and predict vehicle movements [22]. Fake consequential sounds were generated for a 'too quiet' hybrid autonomous car using chords of pure tones with frequency modulation based off the current car's speed, and attaching the speaker directly against the chassis to intentionally generate realistic resonance. People reported preferring interactions with the car that had these added sounds, as they found the intent of the car to change speed or yield was easier to predict.

The above studies help to illustrate the effects that consequential sounds of robots can have within HRI, and thus the importance for researchers and engineers to have an understanding of the variety of consequential sounds produced by different robots, and how to work with these consequential sounds. To further this understanding, consequential sounds generated by a selection of five example robots are herein presented and compared.

6.5.2 Consequential Sounds of Different Robots

A brief description of five different form-factor robots and their consequential sounds is presented below. Footage of these robots demonstrating a large range of different consequential sounds across a 30-45 second window has been supplied as supplementary material www.soundandrobotics.com/ch6, with a visual of each robot included in Figure 6.2.

Go1 Quadruped is a medium sized robot manufactured by Unitree [40] for both research and consumer use. The robot weighs 12 kg unloaded and has similar standing dimensions to a small-medium sized dog at approximately 65cm(L) x 28 cm(W) x 40-46 cm(H). Go1's consequential sounds are predominantly generated by its 12 brushless DC motors with three in each of its 4 legs. These sounds can vary considerably between gaits (e.g. walking, running or stair climbing), and more plyometric movements (jumps, twists and thrusts to stand on its hind legs). The robot has notable consequential sounds from the cooling fans for the multiple on-board computers. In addition to consequential sounds, the robot also produces significant vibrotactile sounds generated from the footpads interacting with ground surfaces.

Pepper is a social humanoid robot created by SoftBank Robotics [29] with a wheeled mobile base, an anthropomorphic upper torso, arms and head as

FIGURE 6.2
Robots as seen in supplementary material videos [1]. From top left to bottom right: Custom multi-rotor drone; Go1 EDU PLUS with 2D Lidar quadruped (Unitree); Pepper Social Robot (SoftBank Robotics); Jackal mobile UGV robot (ClearPath Robotics); Yanshee Humanoid Robot (UBTECH).

well as a tablet-like screen on its chest. Pepper stands 120 cm tall, and weighs approximately 28 kg. Most of Pepper's consequential sounds are of a fairly low intensity as most of the actuators are relatively low power. The Pepper is designed to make expressive motions with the head and arms. Consequential sounds are often produced by the friction between sections of chassis as the robot moves around, with the motors generating a soft electric whir as each motor switches on and off. There are a total of 20 actuators across the arms, head, and wheeled base, with the majority in the arms (6 per arm).

A custom multi-rotor drone (quadcopter) was designed and built for HRI research on developing semi-automated piloting drone software to assist inexperienced pilots [2]. The 'very small' quadcopter weighs 1.8kg and measures 38cm(L) x 38cm(W) x 30cm(H). The four actuators are off-the-shelf DC brushless motors which were designed for continuous spin uncrewed aerial vehicle (UAV) applications. The intensity of the consequential sounds generated by these rotor motors mask any sounds generated by other parts of the drone such as compute, resonance of frame, or mild vibrotactile sounds of the battery against the chassis. The general high intensity of these consequential sounds makes any changes in sound profile from other variables hard to detect.

Jackal UGV is an autonomous mobile outdoor robot created by ClearPath Robotics [28]. It is an entry-level field robotics research platform which weighs 17kg and is approximately 50cm long. Jackal has four identical, large, high-powered motors used to drive its four wheels. As the wheels tend to spin many times in the same direction to locomote the robot, mostly broad spectrum sounds are generated, with acute stop/start sounds as the movements change. Sound properties such as intensity and pitch vary by speed and direction that the robot is moving in.

Yanshee Humanoid Robot is a small, table-top humanoid robot manufactured by UBTECH [39]. It is an open-source platform intended for education and research which weighs 2.05kg and stands approximately 37cm tall. The Yanshee produces high-intensity sound for such a small robot due to its inexpensive DC motors creating ego-motion sounds from the many points of friction. This robot has a large number of actuators (17 servo motors), which are coupled together with aluminum alloy and ABS links that tend to resonate as the motors turn.

6.5.3 Case Study: Comparison of Consequential Sounds across Different Robots

A common method for the analysis of sound is through the use of a spectrogram, which produces a visual representation of the sound frequency content and intensity over time. Figure 6.3 shows a spectrogram for each of the previously described robots. The y axis of a spectrogram represents the frequency, usually between 0 and 20,000 Hertz, to cover the range of human hearing. The x axis of a spectrogram displays the time, and is typically considered in milliseconds or seconds, dependant on the analysis purpose. The color intensity or brightness of the spectrogram represents the intensity or amplitude of the frequency content of the audio signal. Brighter regions indicate higher sound intensity or amplitude, while darker regions indicate lower sound intensity or amplitude.

There are many standard features that can be extracted from a spectrogram. One common feature is the spectral centroid, which measures the center of gravity of the frequency distribution in a signal, providing a measure of the "brightness" or "darkness" of the sound. Two other features include spectral bandwidth and spectral flatness, which provide additional information about the frequency content of a signal. Spectral bandwidth measures the range of frequencies present in a signal, and can provide insight into the "sharpness"

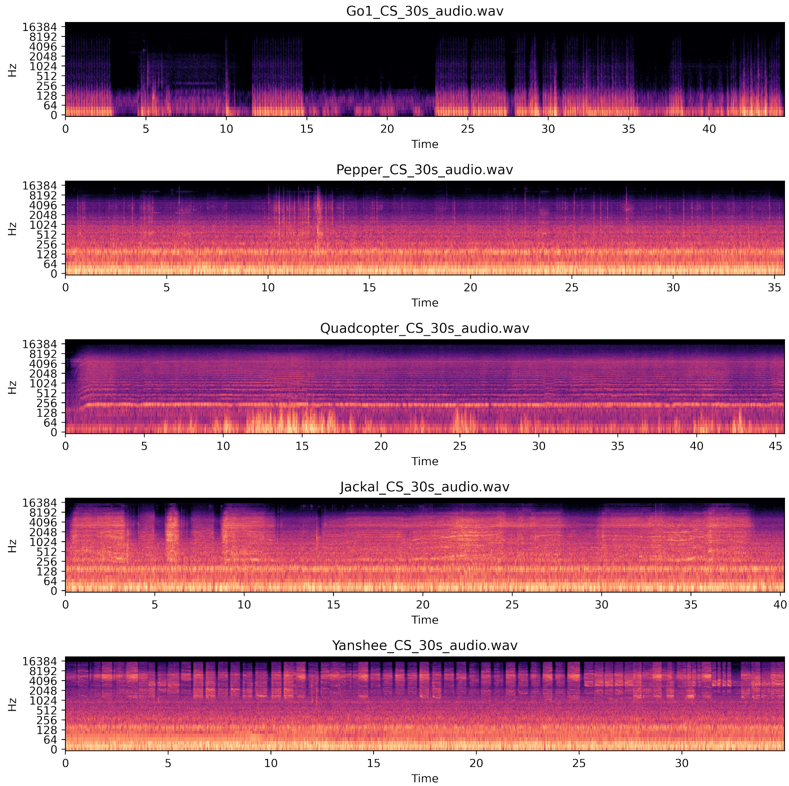

FIGURE 6.3
Spectrograms of the five case study robots.

or “dullness” of the sound. Spectral flatness measures the degree to which a signal’s energy is spread evenly across its frequency spectrum, and can provide insight into the “tonality” or “noisiness” of the sound. Figure 6.4 displays the spectrograms for the case study robots, with overlaid spectral centroid and bandwidth. Whilst there are other methods to visualise and analyze audio, spectrograms are a good start as they can be easily generated, and provide many easily analysable features within an easy to understand visualization. The properties from spectrograms can additionally be used for machine learning analysis and generation.

From a visual analysis comparing the robots’ spectrograms, multiple features of note become clear. The Go1 quadruped, shows clear fluctuations in the spectral centroid with each step the robot takes. This spectrogram also indicates that there are very low noise levels between steps, and that the sound from each step primarily takes place in the lower frequencies. Much of Pepper’s spectrogram comes from underlying low frequencies of ambient room noise,

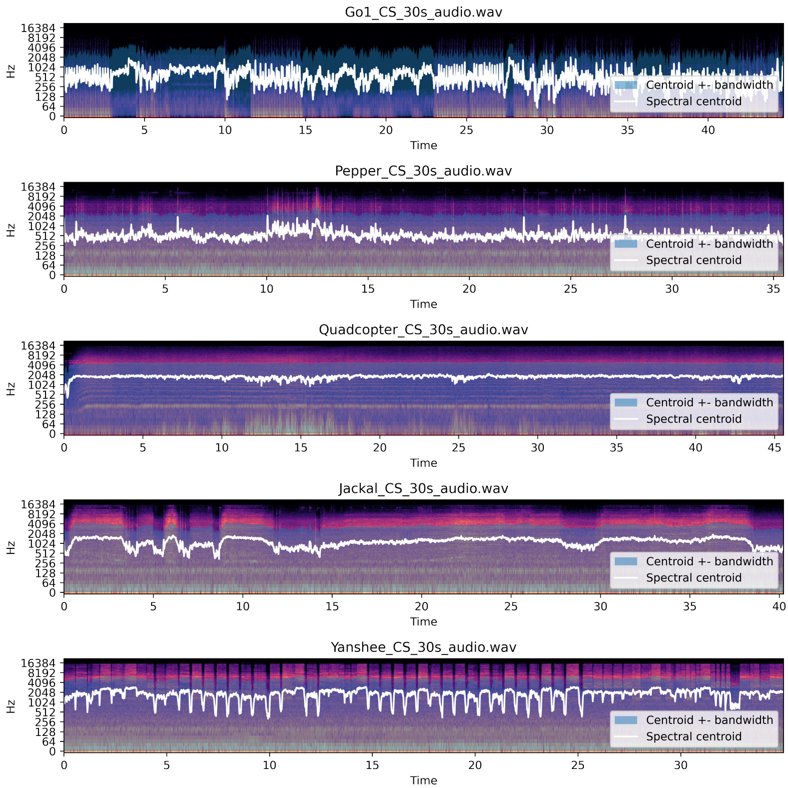

FIGURE 6.4
Spectrograms of the five case study robots: with spectral centroid and spectral bandwidth overlaid.

rather than the consequential sounds themselves. From an audio engineering perspective, it would be common practice to first apply a low-pass filter to remove the lower frequency noise before analysing the robot's consequential sounds. Within boosts of intensity in the higher frequency bands of the spectrogram, very uniform intensity regions can be seen, and with no clear pitches. The quadcopter has a very consistent centroid, with only occasional fluctuations. Of note is the reoccurring lines in the spectrogram, which are heard audibly as the pitch of the hum from the drone. Addressing this reoccurring harmonic series would be important when designing sounds for a drone. The Jackal has the widest, most noise-like signal compared to the other robots, with a more consistent intensity across all frequencies. Audibly, this is perceived as the sound being relatively un-pitched. The Yanshee has a very broad sound across the spectrum, with emphasis on each movement. Importantly for the Yanshee, the higher more piercing sounds can be seen on the spectrogram between 4096 and 8192 Hz, which would be a key consideration for any sound alterations.

Each robot has a unique audio signature in terms of the frequency distribution and intensity of their sound, which becomes more readily apparent through visualization. The analysis of these robots' audio spectrograms provides insight into the characteristics of their consequential sounds, and thus any potential features of these sounds to target for control or alteration.

6.6 Capturing and Altering Consequential Sounds

Given the large range of consequential sounds robots are capable of producing, there are many difficulties and challenges to consider when working with consequential sounds. Whilst the objective existence of consequential sounds can not be changed, there are both hardware and software options available to control or alter the consequential sounds produced. Recommendations for handling the many variables which can impact and change consequential sounds produced by a robot can be found in Table 6.2. Being mindful of these recommendations will help generate reduced and consistent sounds, allowing for accurate capture of consequential sounds, which is necessary prior to applying any further augmentation techniques. Many current industry practices for working with the unintentional sounds generated by a product or machine, may prove useful as a starting point or part-solution for altering consequential sounds in robotics.

To capture consequential sounds, at least one quality microphone must be used to record the sounds. If possible, a condenser microphone with full human audible spectrum range of 20Hz–20kHz should be used, although 60Hz–18KHz should suffice. Many consumer-grade microphones come with built-in filtering software to reduce the recording to only include frequencies of sound common in human speech. As such, these microphones should be avoided as it is more effective to capture the raw sound across the entire human audible spectrum. In most circumstances, it is ideal to place the microphone as close as possible to the sound source (i.e. the robot or actuator) and in such cases an omnidirectional (flat recording spectrum in all directions) microphone is ideal. In cases where the microphone is placed near a camera, a directional or cardioid microphone facing toward the robot may be better, to avoid picking up room ambient sounds. When intending to mount the microphone onto the robot, other criteria should be considered such as minimising the payload. Microphone placement is extremely important and should be selected based off how the user will perceive the sounds, how discreet the microphones should be during the interaction, the types of changes that will be made to the consequential sounds, and how many microphones are available. In general, the microphones should be placed at any point where consequential sounds could be audibly perceived

TABLE 6.2
Recommendations for generating reduced and consistent consequential sounds.

Variable	Recommendations
Component choice (actuator type and specs)	Be mindful of the sounds generated by chosen actuators, choose quieter actuators if possible, but don't sacrifice functional requirements for less sound
Product layout (how components are arranged relative to each other)	Avoid shapes and arrangements which resonate at frequencies which the actuators typically move at
Turning speed of actuators	Speed requirements are mostly fixed by functional movement requirements. Use sound alteration techniques when useful
Full revolutions (continuous spinning) versus smaller angle motions (e.g. precise positioning of stepper motors)	Try to reduce actuator accelerations when starting/stopping actuators such that sound changes are less abrupt
Material choices for inactive components (especially housings/chassis)	Avoid using materials which conduct sound well or that naturally resonate at frequencies which the actuators typically move at
Current power levels	For critical human-perception use-cases (such as experiments and product trials), attempt to maintain close to full charge when possible. For many robots, above 80% power should be ideal, or above 60% for high-drain applications
Temperature of environment and components	Minimise strong environmental temperature changes by running experiments in controlled indoor environments or on days where temperatures are mid-range and fairly consistent
Surfaces standing on/or objects interacting with	This may vary substantially with robot use cases. Sound deadening materials can be attached to contact points such as footpads and grippers
Imperfections in the robot as it ages e.g. dents in chassis	If a robot component begins creating undesirable noises from age, that component should likely be replaced

from, whilst avoiding anything that adds noise to the signal e.g. clipping from microphone being too close to fans or moving parts, or nearby cables causing electrical interference. Common scenarios and suggested microphone locations can be found in Table 6.3. There are also software requirements to record

TABLE 6.3
Recommended microphone placements to record consequential sounds in HRI scenarios.

Scenario	Microphone Placement
Video or online study	At location of person i.e. near camera
Multi-person and general in-person studies	As close as possible to actuators (omni-directional microphone recommended)
Individual customised sound perception (single participant)	Close to the single person
Recording consequential sounds with intent to alter them	As close as possible to actuators (omni-directional microphone recommended)

consequential sounds. Fortunately, most recording software covers the full audible human spectrum range 20Hz–20kHz, and there exists a variety of professional recording software, hobbyist phone apps or Python libraries which should be suitable depending on individual requirements such as budget and onboard versus offline recording requirements. The following techniques can help to improve the effect that consequential sounds can have.

Technique 1. Actuator Choice: Many current solutions focus on hardware design for sound, i.e controlling for consequential sounds by pre-designing actuators and other components to minimise any potentially negative sounds. Whilst useful, this is not feasible to do for every different actuator and robot, may necessitate non-desirable compromises on other functional requirements, and doesn't allow for adaptability for variables which change consequential sounds and their perception. Currently, most off-the-shelf actuators have not considered sound during their design. Therefore, choosing robot actuators that have in-built noise control might necessitate non-desirable compromises on other functional requirements. Whilst it would be possible to control for consequential sounds in the manufacturing process for future or custom actuators, it would still be challenging to predict the full effects of sonic interactions across every robot using a specific actuator across a multitude of contexts. Some existing industry sound reduction techniques (such as dampening) may also add unnecessary weight to the robot, increasing power requirements and potentially putting further strain on the actuators increasing generation of consequential sounds.

Technique 2. Audio Recording and Analysis: A good option to record consequential sounds is to use a Digital Audio Workstation (DAW), which is a piece of software used for recording, editing and producing complex recorded audio. DAWs are typically used by professional sound engineers for music production and sound effect generation, with Ableton, Avid Pro Tools, Logic Pro and Cubase being some of the most popular. A DAW can be used to look at the full frequency spectrum of sounds, and allow identification of good broad spectrum consequential sounds versus any acute/abrupt sounds which might be worth altering. Unless access to a pro-level DAW is readily available, these may not be ideal for consequential sound analysis as they can be expensive, have high learning curves, and may not perform well in real-time on the onboard compute within a robot. For HRI applications, the Audacity DAW is recommended for recording of consequential sounds, as it runs well on typical embedded architectures including Linux/Ubuntu, Intel/ARM processors, Raspberry Pi's and Nvidia Jetsons. In addition to Audacity, recommended tools to analyse consequential sounds for HRI include Sonic Visualiser https://www.sonicvisualiser.org and Librosa for python-based analysis. Producing a spectrogram (as shown in section 6.5.2) can be a good start for analysing consequential sounds.

Technique 3. Masking: A widely accepted psychoacoustic method for hiding any negatively perceived sound is masking [7, 27, 34], A more pleasant sound (masker) is used to reduce the detection of an unpleasant sound (maskee), thus improving the overall perception. It is worth noting that the masker sound does not necessarily have to be loud and add noise, as it can have an effect by containing similar frequencies as the sound to mask, even if it is the same (or slightly lower) intensity. Whilst masking is not often used in current robotics practice, this does have the potential to be applicable to a large range of different robotic platforms, adaptable to specific contexts and customizable to individual preferences. It is known that broad-spectrum sounds tend to work best for masking [12, 27] i.e. non-pure tones covering a wide frequency band of sounds across the entire human audible spectrum, however more research is required on what other properties make good consequential sounds masks. Two particular attributes which require further investigation are specific sounds that most people enjoy and the timing with which to produce these sounds. Regarding noise types to use for masking, standard broad spectrum noises have shown good promise, especially pink and brown (Brownian) noise bands with sound intensities matched to average human loudness perceptions [5, 27]. In terms of timing, it is currently unknown if temporal (slightly before or after the sound), simultaneous (during the sound) or using the masker continuously to feign proxemics is most effective.

Technique 4. Other Software Adjustments: In addition to masking, there are several other promising software techniques to alter consequential sounds in an adaptive way that could be retrofitted to existing robots. Most of these are yet to be applied to consequential sounds in practice. Adaptive techniques are particularly useful when working with consequential sounds as it is hard to tell what the full sound profile of the robot will be during design (i.e. prior to construction) [24]. Another benefit of using many of these techniques is the real-time adaptability to allow for a variety of contexts, and the ability to personalise sound alterations to preferences of individual users or co-inhabitants. To make use of real-time sound augmentation techniques, additional software (to generate sound alterations) and hardware (speaker to emit the generated sounds) may be required. Selection criteria for speakers are similar to requirements for microphones, with speaker being placed as close as possible to sound sources whilst avoiding anything that adds noise to signal e.g. cabling too close to each other, or the microphone feeding the speaker.

Ego-motion sound detection: Existing software based noise mitigation techniques used to improve speech recognition [13, 43] or for contact detection [21] may be useful for handling consequential sounds. In the first case, the real-time changes in noises a robot produces from ego-motion (the friction sounds within motors) can be captured and separated into groups via intensity and localization on the robot using techniques such as Blind Source Separation(BSS) [43]. Noise cancelling techniques to reduce environmental noises within each group can be determined to improve the recorded sound to be parsed for speech recognition. Many newer robots (especially robotic arms and teleoperated robots) are being built with internal accelerometers for measuring ego-vibrations when contacting objects in the environment [21]. Audio processing algorithms have been successfully used to isolate the noises generated by vibrations of the robot's own actuators(ego-vibrations) from contact with the environment sounds(vibrotactile sounds) to increase accuracy in detecting these contacts. Both these techniques work well for removing noise from output sound files, and have potential as the first step of a larger consequential sound solution. There is already significant research on recording and identifying ego-noise from robots [13, 21, 42, 43], however research is scarce regarding the next step, which is using the identified consequential sounds as data to inform other techniques capable of augmenting the audible consequential sounds, which are produced by a robot. For example, this data could be used to inform masking or noise cancellation techniques.

Reducing variance in consequential sounds: Many robots have consequential sounds which alternate between complete silence when stationary, to very acute and high intensity sounds when the robot suddenly begins moving. Both of these are often perceived negatively, leading to people feeling uncomfortable, distracted from a task, or even scared of the robot. One way to alter these consequential sounds could be to add constant sounds to create a sense of proxemics for the robot, giving the robot a consistent passive noise, and making the active motor noises less obtrusive. Other promising methods to minimise noise in less controlled 3D spaces such as outdoors and public spaces are also being researched such as estimating and cancelling time-variant sound in a sphere traveling out from the primary source [20].

Lastly, it is important to review local and federal laws and regulations related to product sound in the jurisdictions the robot will be deployed in. Many places have noise pollution and consumer sound protection laws for maximum sound intensities, and some machines such as autonomous and electric cars also have stipulations on minimum sound volumes.

6.7 Design Implications for Consequential Sounds in HRI

Given the prevalence of consequential sounds in every human–robot interaction, there are notable design decisions regarding how to create successful experiences with robots that either address or negate the impact of consequential sounds. Human perception of consequential sounds can have a notable effect on robot interactions and subsequent acceptance, particularly if these perceptions are not appropriately addressed when robots are deployed into human spaces.

1. Choosing to leave the consequential sounds alone or augment the sounds: An important initial decision for sound designers and researchers is between choosing to cover up consequential sounds, to leave them unaltered, or perhaps alter some sounds and leave others unchanged. Researchers and engineers should aim to identify the relevant sound attributes which may be appropriate or inappropriate for long-term robot use (see section 6.3.1). How consequential sounds might impact human perception, or affect interaction outcomes should be carefully considered. Altering consequential sounds can make the sound more pleasing to people, so any consequential sound alterations should be focused on less desirable sounds without removing perception of the desirable consequential sounds. Leaving some consequential sounds may help to amplify the experience of working with a non-human agent, by allowing people to hear machine-like sounds and thus associate the robot as a mechanical device. This could reduce some false expectations of human-like capabilities in a robot [3]. Whatever the decision, it is important to accurately capture consequential sounds to both test their prevalence, as well as to make effective improvements. Sound alterations that could be considered include softening contrasts between silence and abrupt sounds, and masking of undesirable sounds including sharp sounds and those centered consisting of limited frequencies such as pure tones without harmonics.

2. Early exposure to accurate consequential sounds to aid in long-term adoption: Given the consistent presence of consequential sounds, sound designers and engineers deploying robots into human spaces should aim to maintain the robot's eventual consequential sounds (including decided upon sound alterations) in any research, case studies, marketing or promotional material. While some of these sounds may be reduced, augmented or altered within specific use cases, its important to note that users should at least be well-aware of the expected sounds they are likely to encounter during long-term interactions with a robot. Most sound research is still conducted with the participant wearing headphones or via pre-recorded videos where a de-contextualised fully controllable sound is used. These scenarios do not capture or display consequential sounds faithfully to participants. Assessing user experience of a robot that does not have consequential sounds may unintentionally be producing a biased response, given that the robot may be perceived within a video as ideal for a scenario, meeting all the functional and aesthetic requirements, but once deployed, is abandoned due to the sound profile not being contextually appropriate.

3. Multiple groups of people to accommodate in one setting: There are two groups of people who are affected by robot consequential sounds: people directly using a robot (i.e. those intentionally interacting with the robot) and people who are sharing an environment with a robot (i.e. are colocated within the same space as a robot). Different user groups will experience these sounds in different manners, whether the sound causes notable distraction for people sharing a space with a robot, or the sound becomes part of the interaction experience with the user. Robots in shared spaces may cause disruption to people, or otherwise negatively effect a well-designed robot interaction [38]. A compromise needs to be made to establish a soundscape which minimises negative perception across all concurrent robot stakeholders.

4. Expectations of real robot sounds: When deciding how to alter or present consequential sounds, engineers and researchers should consider what expectations their robot users may already have of the sounds prior to first impression of the robot. Often consequential sounds are stripped from video promotion, or a musical soundtrack is played over the top, leading to users being unaware that robots even make consequential sounds. This could lead to a high probability that expectations will be mis-aligned, thus that the robot does not sound correct on the first encounter. This could contribute to someone becoming uncomfortable, and forming a negative initial association with the robot, as "expectation confirmation theory" [35] has been violated. People may habituate over time [25] and become comfortable with or enjoy these sounds, but this does require that the person persists with interacting with the robot long enough for the sounds to be familiar, which is typically not the case in experiments. When possible, research on sound should be done in person in a real 3D space. This will allow for correctly gauging the effects of the full sound spectrum by considering 3D sound effects, and proxemic effects. If initial tests must be conducted using video clips, then sound tests should be included at the start of any experiment to verify what frequencies and intensities of sound people can hear within the videos. This allows for control of variables such as sound equipment settings, personal sound sensitivity, and hearing capabilities.

6.8 Research Potential for Improving Consequential Sounds

Given the extensive prevalence of consequential sounds in human–robot interaction, this is clearly an important area for future research. Below are several clear opportunities for potential research avenues to explore to further understand and improve how consequential sounds impact robot engagement and perception.

1. Real world or naturalistic lab settings: Experimental verification of results are not often conducted using in-person naturalistic environments. Most existing research on robot sounds uses only pre-recorded footage where the consequential sounds have been completely stripped [24], which is clearly not comparable to a real-world situation. Other research involves masking consequential sounds by playing higher intensity sound directly into participants ears through headphones [9]. However, it is not feasible to have humans augment their own hearing by wearing headphones whenever interacting with a robot in a real setting. Thus research is required on pragmatic solutions to improving consequential sounds directly on the robot itself, in real world or naturalistic lab settings designed to imitate the home, workplace and public spaces.

2. Verification of results on different robots: Due to financial costs in acquiring or accessing robots, as well time costs in familiarization of setting up new robotic platforms, most robotics research is done on a single robot. As robots are known to each produce different consequential sounds, it is likely that findings of useful techniques and preferred sounds could vary between different robots, with what works well on a quadruped, either not transferring to a humanoid or UGV or needing alterations to work successfully on different platforms. Thus further research using multiple robots to verify results would be useful. An ideal circumstance here would be more standardization in research to allow for collaborations where researchers could exchange their setups so other researchers could verify the results using a different robot with which they are already familiar.

3. Development of full solutions for real-time adaptations of consequential sounds: As noted in Section 6.6, there are currently no end-to-end solutions for capturing and augmenting variable consequential sounds in real-time. Research on real-time sound alterations that can change between context and individual personal preferences will be immensely useful once robots are deployed heavily in offices, homes and public spaces. Additionally, if robots are able to control their sounds produced contextually, there is potential to further enhance environments beyond just improving consequential sounds. For example, developed algorithms to alter consequential sounds could be extended to include sounds to improve mood or concentration, target an individual's health conditions or positively augment the consequential sounds of other machines within the shared environment.

4. Larger and multiple-participant studies: In contrast to individual adaptations useful for cases where the robot has a primary interaction target, it would be beneficial to the field to conduct multiple-participant studies to uncover which techniques are effective for groups of people, such as is common in offices or public spaces. In addition, studies with larger numbers of participants would help to identify sound alterations, which are more globally accepted by people.

6.9 Conclusion

Consequential sounds produced by robots is clearly a persistent phenomenon which will continue to occur over time for all robots and all human robot interactions. Whilst there are several part solutions available to address consequential sounds, there is no known full solution for even a single robot, let alone the millions of robots that are or will be collocated with people in the near future. Consequential sounds vary between different robot platforms but also over time with the same robot, so it is important to accurately collect and analyse the sounds specific to each robot and contextual use. There are a variety of techniques which researchers and engineers can use to capture and alter robot consequential sounds in order to improve their perception, and HRI experiences. More research is required to further streamline techniques for consequential sound alterations in order to produce refined techniques that can work in the real world for a variety of different contexts, with different robots, and be personalizable to different people.

Attributions

This book chapter (and associated research) was partly supported by an Australian Government Research Training Program (RTP) Scholarship.

Bibliography

[1] Allen, A., Savery, R., and Robinson, N. Sound and Robotics: Chapter 6 supplementary material, www.soundandrobotics.com/ch6, (2023).

[2] Backman, K., Kulic, D., and Chung, H. Learning to assist drone landings. *IEEE Robotics and Automation Letters 6* (2021), 3192–3199.

[3] Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics 1* (2009), 71–81.

[4] Basner, M., Babisch, W., Davis, A., Brink, M., Clark, C., Janssen, S., and Stansfeld, S. Auditory and non-auditory effects of noise on health, The Lancet (2014), 1325–1332.

[5] Cha, E., Fitter, N. T., Kim, Y., Fong, T., and Mataric, M. J. Effects of robot sound on auditory localization in human-robot collaboration. *ACM/IEEE International Conference on Human-Robot Interaction* (2018), 434–442.

[6] VAN EGMOND, R. The experience of product sounds, Product Experience (2007), 69–89.

[7] Fastl, H., and Zwicker, E. Masking, Psychoacoustics: Facts and Models, (2007), 61–110.

[8] Fastl, H., and Zwicker, E. Sharpness and sensory pleasantness, Psychoacoustics: Facts and Models, (2007), 239–246.

[9] Frederiksen, M. R., and Stoey, K. Augmenting the audio-based expression modality of a non-affective robot. In *2019 8th International Conference on Affective Computing and Intelligent Interaction, ACII 2019* (2019).

[10] Frederiksen, M. R., and Stoy, K. Robots can defuse high-intensity conflict situations, In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IROS2020. pp. 11376–11382.

[11] Frid, E., Bresin, R., and Alexanderson, S. Perception of mechanical sounds inherent to expressive gestures of a NAO robot - implications for movement sonification of humanoids. In *Proceedings of the 15th Sound and Music Computing Conference: Sonic Crossings, SMC 2018* (2018), pp. 43–51.

[12] Hall, J. L. Auditory psychophysics for coding applications, Digital Signal Processing Handbook (1999), Chapter 39.

[13] Ince, G., Nakadai, K., Rodemann, T., Tsujino, H., and Imura, J. I. Whole body motion noise cancellation of a robot for improved automatic speech recognition. *Advanced Robotics 25* (2011), 1405–1426.

[14] Jariwala, H. J., Syed, H. S., Pandya, M. J., and Gajera, Y. M. Noise pollution & human health: A review Proceedings of Noise and Air Pollution: Challenges and Opportunities (2017).

[15] Jokisch, O., and Fischer, D. Drone sounds and environmental signals – a first review. Proceedings of ESSV2019_84 Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung (2019), pp. 212–220.

[16] Jouaiti, M., and Henaff, P. The sound of actuators: Disturbance in human - robot interactions? In *2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics, ICDL-EpiRob 2019* (2019), pp. 75–80.

[17] Kloet, N., Watkins, S., Wang, X., Prudden, S., Clothier, R., and Palmer, J. L. Drone on: A preliminary investigation of the acoustic impact of unmanned aircraft systems (UAS). In *24th International Congress on Sound and Vibration, ICSV 2017* (2017).

[18] Langeveld, L., van Egmond, R., Jansen, R., and Özcan, E. Product sound design: Intentional and consequential sounds, Advances in Industrial Design Engineering, (2013), 47–73.

[19] Lyon, R. H. Product sound quality - from perception to design. *Sound and Vibration 37* (2003), 18–22.

[20] Ma, F., Zhang, W., and Abhayapala, T. D. Active control of outgoing noise fields in rooms. *The Journal of the Acoustical Society of America 144* (2018), 1589–1599.

[21] McMahan, W., and Kuchenbecker, K. J. Spectral subtraction of robot motion noise for improved event detection in tactile acceleration signals. , In *Proceedings of EuroHaptics*, (2012), 326–337.

[22] Moore, D., Currano, R., and Sirkin, D. Sound decisions: How synthetic motor sounds improve autonomous vehicle-pedestrian interactions. , In *Proceedings - 12th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI* 2020, 94–103.

[23] Moore, D., and Ju, W. Sound as implicit influence on human-robot interactions. In *ACM/IEEE International Conference on Human-Robot Interaction* (2018), pp. 311–312.

[24] Moore, D., Tennent, H., Martelaro, N., and Ju, W. Making noise intentional: A study of servo sound perception. In *ACM/IEEE International Conference on Human-Robot Interaction Part F1271* (2017), pp. 12–21.

[25] Mutschler, I., Wieckhorst, B., Speck, O., Schulze-Bonhage, A., Hennig, J., Seifritz, E., and Ball, T. Time scales of auditory habituation in the amygdala and cerebral cortex. *Cerebral Cortex 20* (11), (2010), 2531–2539.

[26] Device Orchestra. We will rock you on 7 electric devices, https://www.youtube.com/watch?v=Hh9pm9yjmLs, 2019.

[27] Richards, V. M., and Neff, D. L. Cuing effects for informational masking. *The Journal of the Acoustical Society of America 115* (2004), 289–300.

[28] Clearpath Robotics. Jackal UGV, https://clearpathrobotics.com/jackal-small-unmanned-ground-vehicle/.

[29] Softbank Robotics. Pepper the humanoid and programmable robot, https://www.aldebaran.com/en/pepper

[30] Samarasinghe, P. N., Zhang, W., and Abhayapala, T. D. Recent advances in active noise control inside automobile cabins: Toward quieter cars. *IEEE Signal Processing Magazine 33* (2016), 61–73.

[31] Schütte, M., Marks, A., Wenning, E., and Griefahn, B. The development of the noise sensitivity questionnaire. *Noise and Health 9* (2007), 15–24.

[32] Sneddon, M., Pearsons, K., and Fidell, S. Laboratory study of the noticeability and annoyance of low signal-to-noise ratio sounds. *Noise Control Engineering Journal 51* (2003), 300–305.

[33] Song, S., and Yamada, S. Expressing emotions through color, sound, and vibration with an appearance-constrained social robot. In *ACM/IEEE International Conference on Human-Robot Interaction Part F1271* (2017), pp. 2–11.

[34] Tanner, W. P. What is masking? *Journal of the Acoustical Society of America 30* (1958), 919–921.

[35] Taylor, J. L., and Durand, R. M. Effect of expectation and disconfirmation on postexposure product evaluations: A casual path analysis. *Psychological Reports 45* (1979), 803–810.

[36] Tennent, H., Moore, D., Jung, M., and Ju, W. Good vibrations: How consequential sounds affect perception of robotic arms. In *RO-MAN 2017 - 26th IEEE International Symposium on Robot and Human Interactive Communication*, (2017), 928–935.

[37] Tian, L., Carreno-medrano, P., Allen, A., Sumartojo, S., Mintrom, M., Coronado, E., Venture, G., Croft, E., and Kulic, D. Redesigning human-robot interaction in response to robot failures : a participatory design methodology, In *CHI '21: ACM CHI Conference on Human Factors in Computing Systems*, May 08–13, (2021).

[38] Trovato, G., Paredes, R., Balvin, J., Cuellar, F., Thomsen, N. B., Bech, S., and Tan, Z. H. The sound or silence: Investigating the influence of robot noise on proxemics. In *RO-MAN 2018 - 27th IEEE International Symposium on Robot and Human Interactive Communication* (2018), pp. 713–718.

[39] UBTECH. Yanshee website, http://global.ubtechedu.com/global/pro_view-4.html.

[40] Unitree. Go1 - unitree, https://www.unitree.com/en/go1.

[41] De Paiva Vianna, K., Alves Cardoso, M., and Rodrigues, R. M. C. Noise pollution and annoyance: An urban soundscapes study. *Noise and Health 17* (2015), 125–133.

[42] Pico Villalpando, A., Schillaci, G., Hafner, V. V., and Guzmán, B. L. Ego-noise predictions for echolocation in wheeled robots, In *Artificial Life Conference (ALIFE)*, (2019), pp. 567–573.

[43] Wang, L., and Cavallaro, A. Ear in the sky: Ego-noise reduction for auditory micro aerial vehicles. In *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance*, AVSS (2016), pp. 152–158.

[44] Wang, Z., Hu, Z., Rohles, B., Ljungblad, S., Koenig, V., and Fjeld, M. The effects of natural sounds and proxemic distances on the perception of a noisy domestic flying robot. *ACM Transactions on Human-Robot Interaction* (2023).

[45] Yilmazyildiz, S., Read, R., Belpeame, T., and Verhelst, W. Review of semantic-free utterances in social human-robot interaction. *International Journal of Human-Computer Interaction 32* (2016), 63–85.

[46] Zhang, B. J., Peterson, K., Sanchez, C. A., and Fitter, N. T. Exploring consequential robot sound: Should we make robots quiet and kawaii-et? In *IEEE International Conference on Intelligent Robots and Systems (IROS)* (2021), pp. 3056–3062.