1

The Landscape of Sound and Robotics

Richard Savery and Trinity Melder

Sound is arguably the biggest missed opportunity in human–robot interaction. While much attention has been given to the visual and physical aspects of robots, such as their movements and appearance, sound has often been overlooked or deemed less important. However, sound can play a crucial role in how robots are perceived and interacted with by humans, especially in social and emotional contexts.

This book emphasizes the ways in which robots can use sound to communicate with humans. This can be through speech, or non-verbal audio, which encompasses all sounds that are created for communication without semantics. We additionally look at ways robots can act as musicians and how musical processes can inform robot design and interaction. This opening chapter presents a brief overview of existing approaches to sound and robotics and offers a summary of the book.

1.1 Overview of Sound and Robotics

To begin this chapter, we conducted a short survey of the use of sound output in robots. We analyzed IEEE's guide to robots (https://robots.ieee.org/), which

DOI: 10.1201/9781003320470-1

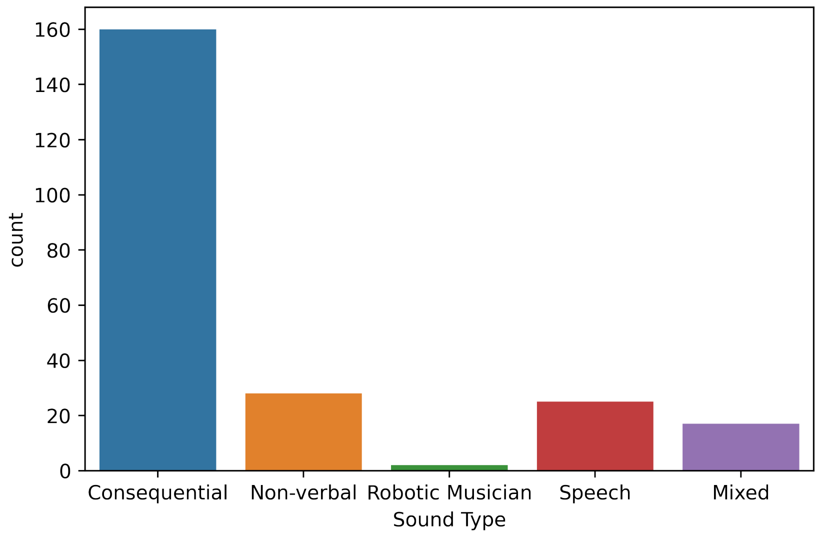

FIGURE 1.1
Sound output from common robot platforms.

includes many common robot platforms and is used in education programs worldwide, making it a valuable resource for our survey. While it is not an exhaustive list, we believe it provides a general overview of the most widely used robots. Our primary goal in the survey was to gain a broad understanding of how sound is currently functioning in robotics, serving as a foundation for the book's focus on leading current research in the use of sound. By examining the robots in the IEEE guide, we were able to identify various broad trends in the use of sound.

Figure 1.1 shows an overview of the types of sound output from the robots in the list. These are broken into speech, non-verbal audio, consequential sound, and mixed. We labeled any robot that primarily used spoken language to communicate as speech, while robots that focused on non-verbal sounds, even if using occasional words, were classified as non-verbal. Mixed robots used speech and non-verbal audio but did not emphasize the use of one or the other. As expected the majority of robots in the list only had consequential sound, which are sounds produced unintentionally as a result of the robot's actions, such as motor or movement sounds. Many of the robots on the list are not focused primarily on human–robot interaction, so a majority focus on consequential sounds is to be expected.

While the survey itself did not show any significant findings, the process made clear many of the existing issues in the research of sound and robotics. The following sections describe some of these key issues and opportunities for future work.

1.1.1 Designers Shift Responsibility

One of the biggest issues in robotics for sound is that many platforms include a speaker, and then leave the responsibility of careful sound design on developers who use the platform. This contrasts the carefully constructed visual and mechanical appearance, which can rarely be changed by end users.

Analyzing the sound between platforms also becomes essentially impossible in these scenarios, as there is no standard sound for a robot, leading to no baseline for comparison. Instead, common robot platforms can be seen with many different sound outputs. While this is a reasonable approach for research, this leaves the field of sound and robotics unable to easily compare results between different robots.

1.1.2 Robot Documentation

Symptomatic of broader robot trends, we had difficulty finding information on some robots that were no longer being produced or discontinued. Often the official website or documentation was unavailable, and we were required to turn to user guides or third-party sources. Likewise, some newer robot platforms described plans for future documentation that were currently unavailable.

This lack of documentation is heightened even further when considering sound and robotics. It was very common to find lengthy descriptions of the design process for the visual appearance, and nothing about the sound. Often robots appearance and motions were demonstrated in videos and publications, while avoiding mention of sound and applying a default speech system.

1.1.3 Categorization

The categories speech, non-verbal audio, and robotic musician are by nature extremely broad. Our original goal included further breakdowns of each category, such as gender for speech, or whether the sounds were emotion-driven, or if they were synthesized or sampled. Ultimately however, it became apparent that it was just not possible to clearly determine how many robots are using sound. This reiterates how much further work is required to understand the role and promote the potential applications of sound for robotics.

1.2 About the Book

Considering the range of sound outputs from robots, this book is split into speech, non-verbal audio, and robotic musicianship. As previously described, the boundary between these categories is not always clear-cut. Nevertheless, the chapters throughout this book are split into sections, with authors primarily addressing one of the forms of communication.

1.2.1 Robot Speech

The first group of chapters in the book delves into the topic of robot speech, exploring the effects of different elements of speech on the user's experience and perception of the robot. Chapter 2 examines the effects of the number of voices and voice type on the storytelling experience and robot perception. The chapter presents findings from an online study comparing human and synthetic voices and the use of single or multiple voices in storytelling. The results suggest that a single voice is preferable for both human and synthetic voices when focusing on the storytelling experience, while the use of different voices is only recommended for synthetic voices when illustrating different characters.

Chapter 3 focuses on how research on vocalizations in human-human interactions can inform the conceptualization of robot sound in human–robot interactions. The chapter highlights three main lessons learned from six examples of human and robot interactions, which include the semantically under-specified nature of both human vocalizations and robot sound, the embodied nature of human sound production, and the need to analyze and design robot sound multi-modally.

Chapter 4 examines the use of speech in social robots for loss-of-trust mitigation. The chapter presents data from two experiments that evaluated the impact of a robot's ability to speak on the user's perception of its trustworthiness, likeability, animacy, and perceived intelligence. The findings suggest that the ability to speak can mitigate the degradation of trust caused by faulty robot behavior, as it increases the perceived intelligence and capability of the robot and makes it appear more like a sentient or living being.

Finally, Chapter 5 explores the topic of grounding spoken language, which involves referencing concepts in the physical world using natural language processing in a robot's environment. The chapter presents a case study of learning grounded language from speech and examines the need for complex perceptual data, the ability to learn to interact directly from speech, and the challenges of learning grounded language from multimodal data. This chapter presents important directions and considerations that are widely overlooked when applying speech to robot systems.

1.2.2 Non-Verbal Audio

The second section addresses the wide range of sound from robots that is non-verbal. The section begins with Chapter 6 which focuses on consequential sounds; the non-intentional sounds from robots such as motor noise. The chapter describes positive and negative effects these sounds can have on interaction, and then highlights the research gap in sound design for consequential sounds and suggests techniques for improving human–robot interaction success.

Chapter 7 explores how a robot's auditory communication can be enhanced by emitting sound across loudspeakers in distributed audio environments. The chapter discusses the design themes for applying interactive sound installation techniques in the context of human–robot interaction, and proposes a generalized design framework for spatial robot sound that can be broken down into three key locations – the robot, objects of interest in the environment, and the space itself.

Chapter 8 examines the concept of sonification and its application in the context of autonomous cars. The chapter presents four different approaches to sonification, discussing the benefits, challenges, and future directions of this emerging field. The goal of the chapter is to describe divergent approaches and cover a range of future possibilities when notifying robot data.

Chapter 9 of this book discusses the use of nonverbal sounds as a means of communication between humans and robots. The authors propose a low-dimensional parameterized communication model based on nonverbal sounds, which is validated using an online survey to investigate the users' experience and shared task performance. The study found that simplified NVS communication could facilitate human–robot collaboration, leading to positive user experience, increased interest in subsequent interactions, and increased collaborative performance.

Chapter 10 describes emotional musical prosody and explores the relationship between personality traits and emotional responses in music-driven emotional interaction in robotics. The authors focus on two of the "Big Five" personality traits, Neuroticism and Extraversion, and investigate how varying the degree of emotional response through sound can be used in robotic systems to demonstrate different personality types. The study found that all human personalities prefer to interact with robots showing low Neuroticism and high Extraversion emotional responses over the short term.

Chapter 11 again uses emotional musical prosody, however focuses on robotic groups. The authors conducted two studies to analyze the impact of emotional musical prosody on trust, likeability, perceived intelligence, and emotional contagion among the participants. The findings suggest that sound and music have broad potential use cases in group human–robot communication and can shape participant responses to robots.

1.2.3 Musical Robots and Robotic Musicians

The final section of the book includes chapters focusing on the intersection of music and robotics. Chapter 12 presents an overview of the field of robotic musicianship, and discusses the evaluation of these systems. This chapter also frames the roles robotic musicians can have in exploring the social and emotional elements of human–robot interaction.

Chapter 13 explores the portrayal of robotic musical emotions through sound and gesture in three different platforms: Stretch by Hello Robot, Panda by Franka Emica, and SeekerBot, an internally developed social robot. The chapter also discusses the use of the framework to create robotic dance based on musical features from a song, where a 7DoF robotic arm was used as the non-humanoid robot.

The final chapter in the book explores the controversy surrounding posthumous holographic performances of deceased singers in the US and Japan. It analyzes how the reception of AI Misora Hibari differed from that of Western counterparts and highlights the importance of voice and culturally specific contexts in robotics research. Importantly it emphasizes broader issues of the role of music and sound in robotic research.

1.3 Conclusion

The chapters in this book provide a diverse and comprehensive exploration of the intersection of sound and robotics. By examining the relationships between sound and robotics from different angles, this book aims to stimulate further research and discussion on the topic, and to foster a deeper understanding of the role and potential of sound in shaping the future of robotics.