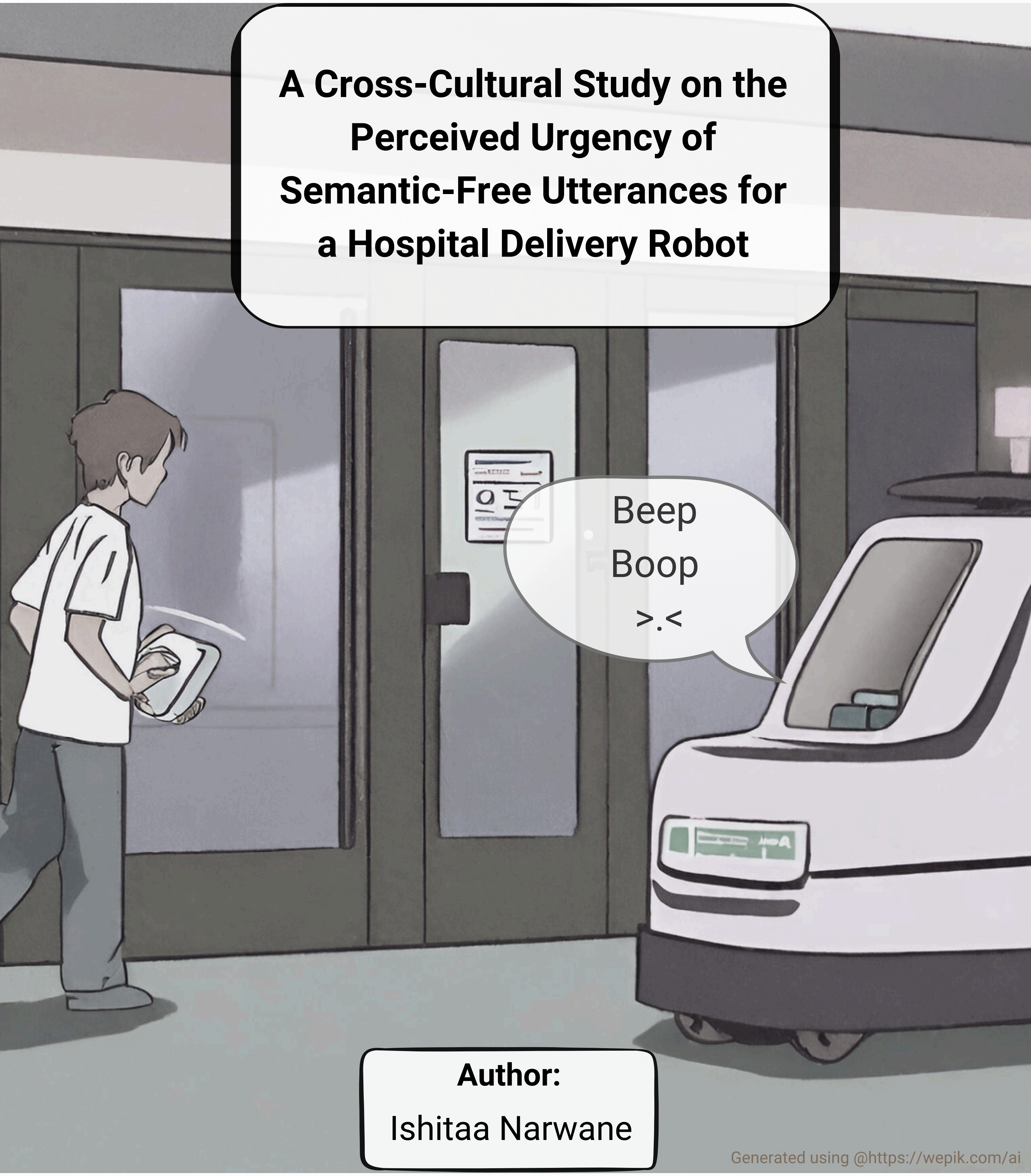
MASTERS' THESIS

A Cross-Cultural Study on the Perceived Urgency of Semantic-Free Utterances for a Hospital Delivery Robot



Author:
Ishitaa Narwane

UNIVERSITY OF TWENTE.

UNIVERSITY OF TWENTE.

Masters Thesis

A Cross-Cultural Study on the Perceived Urgency of Semantic-Free Utterances for a Hospital Delivery Robot

Author:

Ishitaa Narwane

MSc Interaction Technology

August, 2024

Supervisors Committee:

dr. Khiet Truong (Chair).

Hideki Garcia Goo (Co-Chair)

dr. Armağan Karahanoğlu (External)

Faculty of Electrical Engineering,
Mathematics & Computer Science,
University of Twente.

Abstract

With social robots becoming ubiquitous, like smartphones, and their presence in sectors like healthcare increasing, effective communication strategies for such public environments with diverse users are essential. Natural language is common in Human-Robot Interaction (HRI), which however faces limitations, especially in cross-cultural contexts due to inconsistency, complexity, and bias in training data.

Semantic-free utterances (SFUs), which are vocalizations devoid of semantic content, emerge as a promising solution. SFUs, used by media droids like R2D2, WALL-E, and Minions, convey emotion and intent without language barriers, and their ease of implementation provides designers with greater control compared to NLP.

This study examines the potential of SFUs for signalling urgency in a hospital delivery robot, focusing on Indian and Dutch national cultural backgrounds. Using Gibberish Speech (GS) and Non-Linguistic Utterances (NLUs) from the EMOGIB and BEST databases, which convey anger, participants' perceptions of these SFUs were assessed for urgency through an online survey in two parts. Part 1 (Pre-Study) narrowed down the SFUs and investigated the correlation between anger and dominance, and Part 2 (Main-Study) assessed selected SFUs for urgency.

Results showed no significant main effects for culture or database alone, but a significant interaction between them. Qualitative feedback indicated that pitch, tone, and loudness influence urgency perception across cultures. A correlation between Dominance and Urgency was also found.

These findings support SFUs' effectiveness in conveying urgency amongst two contrasting cultures, with implications for designing communication strategies for service robots in multicultural environments.

Acknowledgment

With heartfelt gratitude, I present this thesis, reflecting on a journey that was a roller-coaster ride filled with unexpected twists and turns. I owe immense thanks to those who supported me at every corner, making this journey both bearable and incredibly rewarding.

First and foremost, my deepest gratitude goes to my thesis supervisor, Prof. Dr. Khiet Truong, and co-supervisor, Hideki Gracia Goo. Your expertise, unwavering support, and patience transformed what could have been a challenging path into an exhilarating adventure. From the bottom of my heart, thank you. I want to express my sincere gratitude to Prof. Dr. Armağan Karahanoglu for joining the committee and providing invaluable feedback as an external member. My study advisor, Eric Bong, guided meetings when I encountered hurdles and ensured a smooth path forward. I also want to thank Prof. Dr. Dennis Reidsma, the godfather of I-Tech, for his approachability and guidance during this master's journey. Prof. Dr. Edwin Dertien, thank you for your course on social robot design, which introduced me to social robotics and sparked my interest in HRI research. Prof. Dr. Christina Zaga, your teachings on human-centred design and social justice in AI greatly impacted my personal and professional life. Thank you to all the professors.

To my parents, your love and encouragement were the safety harnesses that kept me secure on this ride. This accomplishment is as much yours as it is mine. Thank you for being my constant source of strength, Aai and Baba, my best friends. I am deeply grateful to my Uncle, Aunt, and Nimish, whose love, support, and warmth provided me with a comforting home away from home. Your presence made this journey easier and more comforting. To my sister Rucha and brother-in-law Ketan, thank you for being my weekend escape, helping me recharge and find balance. Kushal, thank you for always being there, and William, for your insightful guidance. Meenal, your unwavering support has been a great source of comfort and strength. Yadhu, Tufail, Karan, Abhay, Vishakha, Nina, Vinay, Richa, Abhishek, and Nikhil, thank you for making my time in Enschede so memorable. Mayuresh and Snigdha, your guidance and encouragement have been invaluable. Mohit and Pallavi, thank you for your virtual support, which has meant the world to me. If I have unintentionally left anyone out, please forgive me—writing this is a bit overwhelming Grateful to everyone for filling this roller-coaster ride with joy and laughter and for keeping me sane. Your love and support made every challenge easier to face and every success more meaningful.

Last but not least, I offer my thanks to the grace of the Almighty, truly grateful.

Contents

List of acronyms

HRI Human-Robot Interaction . . . vi

HHI Human-Human Interaction . . . x

H-HRI Healthcare Human-Robot Interaction . . . 2

cHRI Child Human-Robot Interaction . . . 17

NLP Natural Language Processing . . . 2

SFUs Semantic-Free Utterances . . . vii

GS Gibberish Speech . . . 2

NLUs Non-Linguistic Utterances . . . 3

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

As technological advancements continue to shape the world we live in the integration of service-providing robots is becoming increasingly prevalent in various industries and our daily routines. In 2021 alone, over 121,000 units of service robots were sold, indicating a growing trend in their adoption[1]. Their presence now extends beyond industries to healthcare,[2] airports, and homes [3].

Given the growing elderly population and impending shortage of healthcare workers[3], the role of robots in healthcare is becoming increasingly significant[4]. The global market for robotics in healthcare is projected to expand significantly, increasing from $8,349 million in 2022 to $36,670.6 million by 2033, at a CAGR of 14.4% from 2023 to 2033.[5] As robots become more ubiquitous and integrated into our daily lives, achieving social acceptance of this technology is essential for their seamless incorporation into society [4], [5].

Previous research indicates that an individual's cultural background, exposure to robots through media, and personal encounters significantly influence their attitudes towards robots [5]–[7]. Notable variability has been observed in user expectations, perceptions, and acceptance of robots due to one's national culture, particularly when explored through the famous dichotomy of East versus West [5], [8], [9]. Multiple studies in HRI have focused on the cultural background as the national culture, encompassing the values, norms, and practices adopted by individuals who are born and raised in a specific country [5], [8], [10]

[1] https://ifr.org/ifr-press-releases/news/sales-of-robots-for-the-service-sector-grew-by-37-worldwide

[2] https://www.agvnetwork.com/hospital-robots

[3] https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

[4] https://www.who.int/health-topics/health-workforce#tab=tab_1

[5] https://www.marketstatsville.com/robotics-in-healthcare-market?trk=article-ssr-frontend-pulse_little-text-block

As previous research indicates, cultural background significantly influences perceptions and attitudes towards robots [5], [8], [9], it is imperative to evaluate robots' communication in a cross-culture setting to identify if potential differences persist due to one's culture. This emphasizes the importance of considering cultural perspectives when designing robots intended for global markets for successful adaptation, acceptance, ease, and intuitiveness in Human-Robot Interaction (HRI) [5].

In addition to cultural considerations, robots interact with humans through various modalities such as visual, tactile, and auditory channels [11]. Among these, sound or audio communication plays a key role in shaping users' perceptions and interactions with robots, allowing users to make assumptions about a robot's capabilities from a distance [12]. Research in audio communication in HRI often focuses on speech or natural language processing (Natural Language Processing (NLP)), which has its limitations [13], [14]. Moreover, not all HRI applications require natural language communication. Some contexts may require the robot to sound machine-like or convey intent without engaging in back-and-forth communication [13], [14]. More details on this are provided in Section 2.3.3 of Chapter 2).

One promising auditory communication method that is gaining the attention of researchers in HRIis Semantic-Free Utterances (SFUs) [13]. These are vocalizations that lack semantic content but can effectively communicate expression and emotion [13]. SFUs have been utilized and accepted worldwide in media robots, such as the famous droid R2D2 and Wall-E, as well as in commercial robots such as My Keepon, RoboQuad, Nao, and Aibo [13], [15]. It has the potential to overcome language barriers and reduce interaction breakdowns [13], [15]–[18]. However, their application especially in public settings accommodating individuals from diverse cultural backgrounds like hospitals, remains unexplored [17], [18].

1.2 Research Gap

Gonzalez et al. [19] conducted a cross-cultural study using Gibberish Speech (Gibberish Speech (GS)), a type of SFUs (more details in Section 2.3.2), with a sample size limitation of $n = 13$. Despite this limitation, the study emphasized the importance of considering cultural differences in HRI. SFUs have the potential in cross-cultural settings due to being independent of language semantics and widely accepted. However, their use in cross-cultural studies within HRI remains under-explored.

To the best of the researcher's knowledge, the use of SFUs in hospital settings within HRI or Healthcare Human-Robot Interaction Healthcare Human-Robot Interaction (H-HRI) has not been explored. SFUs have the potential to effectively communicate intents such as urgency for hospital robots since efficient task execution by robots is crucial in complex and diverse hospital environments due to the critical and time-sensitive nature of operations. Addressing this gap will enable robot designers to use SFUs in multicultural, diverse HRI settings to communicate intent effectively. This study aims to address the gap in utilizing SFUs in robots for conveying urgency, particularly in hospital settings, which is the main objective of this research (refer to Figure 1.1).

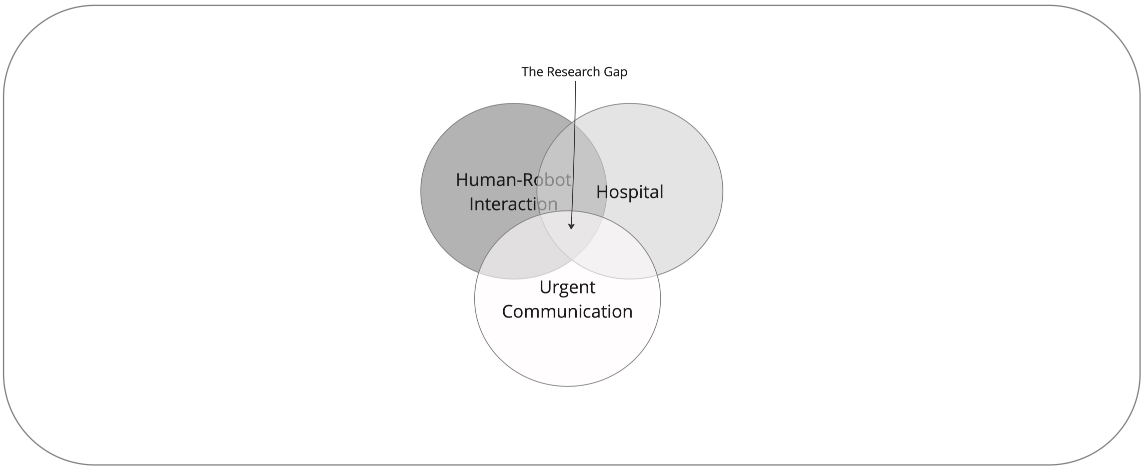


Figure 1.1: Research Gap

Although Non-Linguistic Utterances (Non-Linguistic Utterances (NLUs)), another type of SFUs (refer 2.3.2, have been utilized to convey urgency in auditory warnings and HRI [20], they often resemble alarms or sirens, potentially leading to ambiguities. These ‘alarm-like’ sounds can create confusion by overlapping with ambulance sirens, fire alarms, or monitoring device alarms, particularly in noisy hospital environments where critical alarms need immediate attention. This cacophony of alarms and other noises can negatively impact hospital staff and patients, potentially leading to ‘Alarm-Fatigue’, caused by sounds in the hospital, as discussed by Özcan et al. [21], [22].

Thus, this study explores the use of SFUs to mitigate these challenges and to improve communication efficiency for robots in critical contexts like hospitals, addressing difficulties such as alarm fatigue and ambiguity. This research aims to improve human-robot collaboration in diverse and multicultural hospital environments by examining how SFUs can effectively convey urgency across different cultural contexts, resulting in more intuitive, inclusive, and universally understood across different languages and cultural settings effective sound.

1.3 Goal of this Study

The goal of this study is to explore the application of (SFUs) for communication of Urgency for a hospital delivery robot Harmony in Figure 1.2 which is responsible for bio-sample deliveries. Given the criticality of tasks at a hospital, effectively conveying urgency to nearby individuals becomes essential to prevent human interference that could delay task completion and potentially impact patient outcomes.

Figure 1.2: Harmony Robot used in this study [1]

Cultural background significantly influences perceptions of robot communication [8]. Individualism and collectivism are important characteristics in understanding how societies perceive the between individuals and groups [23], [24]. Individualistic cultures, such as the Netherlands, individualism i.e., values personal autonomy, whereas collectivist cultures, such as India, value group welfare [23]. These differences influence cognitive processes, and psychological experiences [23], and hence could influence the interpretation of Urgency HRI. As a result, evaluating cultural differences can lead to more effective design strategies for robot communication across diverse user market for robots, globally.

Through a cross-cultural study, this research seeks to investigate how two types ofSFUs, Gibberish Speech (GS) and Non-Linguistic Utterances (NLUs), from two different corpora, BEST (corpus for NLUs) and EMOGIB (corpus for GS), influences the perception of Urgency. The study examines how national cultural differences in beliefs, practices, and social norms between participants from India (a collectivist culture) and the Netherlands (an individualistic culture) shape their perception of Urgency communicated by robot GS and NLUs [8], [9]. Additionally, it explores whether perception differs between the two types of SFUs.

It study was conducted in two phases. The pre-study (Chapter 4) for evaluating the perception of these SFUs for Dominance and Anger. The second phase, the main-study (Chapter 5), focused on evaluating SFUs on the perception of conveying Urgency. To investigate this, the following research questions were formulated:

> *(RQ:) What is the effect of the type of Dominant Semantic-Free Utterances from two different corpora (BEST vs. EMOGIB) on individuals from two distinct cultural backgrounds (Indian vs. Dutch) when these utterances are utilized by a hospital delivery robot?*

This main research question also leads to the formulation of the sub-research question:

SRQ_1 (Database Focused): What is the effect of dominant SFUs from two different corpora (EMOGIB vs. BEST) on the perception of urgency in robot behaviour?

SRQ_2 (Culture Focused): What is the effect of dominant SFUs in robot behaviour on the perception of urgency across Dutch and Indian cultural backgrounds?

The study was conducted online using the survey platform Qualtrics. Python packages (pandas, numpy, matplotlib, scipy) and R packages (ARTool, emmeans, multcomp, rcompanion, ggplot2, psych) are used for data analysis. More details on the hypothesis are in the section Research Objectives in Chapter Databases and Research Objectives 3 and on data analysis are in Chapter Pre-study 4 and Chapter 5.

1.4 Report Outline

Chapter 2 comprehensive review of the existing literature on SFUs, urgency, anger, dominance and rudeness perception in cross-culture, as well as, HRI. Chapter 3 provides details on the corpora from which the SFUs will be utilized for evaluation, and outlines the research objectives and hypotheses. The study is divided into two phases: a pre-study phase and a main-study phase. Chapter 4 focuses on the pre-study, covering methodology, analysis, results, and discussions. Chapter 5 addresses the main-study, focusing on the analysis of the Research Question, methodology, analysis, and results. The findings are discussed in Chapter 6 and finally, concluding remarks with HRI implications of this thesis, future recommendations and limitations encountered in this study are discussed in Chapter 7. It is important to note that the terms corpora and databases are used interchangeably in this context, both referring to the same concept of Database of SFUs.

Chapter 2

Background

This chapter provides the background for the thesis. First the relevant literature in Human-Robot Interaction (HRI) was examined (Figure2.1). In cases where relevant literature was unavailable, the review extended to Human-Human Interaction (HHI), particularly focusing on behaviours and expressions (refer to Figure 2.2). The chapter also examines a form of Human-Robot Communication known as Semantic-Free Utterances (SFUs) in HRI, discussing their classification, potential applications, and practical implications, particularly it addresses the insufficient utilization of SFUs in cross-cultural public settings. Furthermore, it explores the concepts of urgency, anger, and dominance and their roles in effective communication within hospital logistics robots, ensuring these signals are perceived appropriately without being seen as rude.

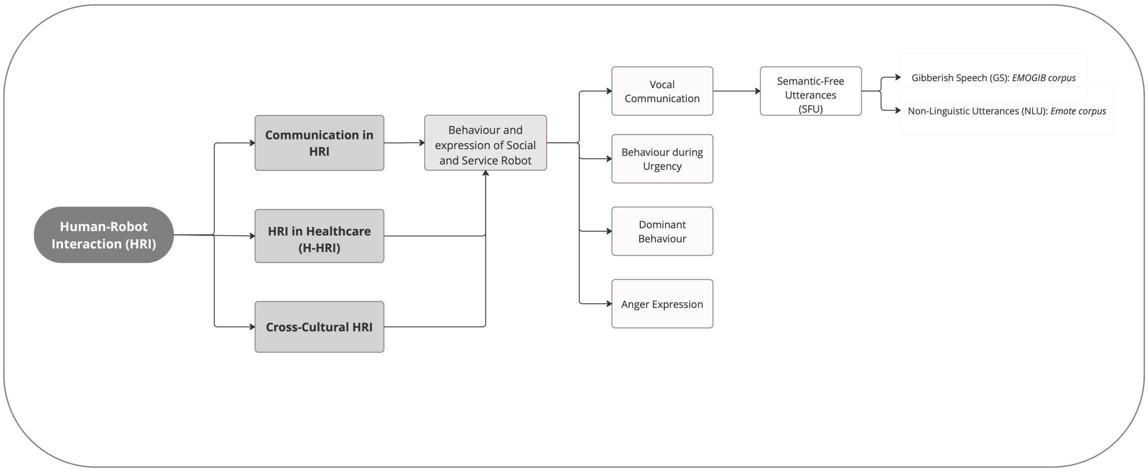


Figure 2.1: Approach of Literature Review for HRI

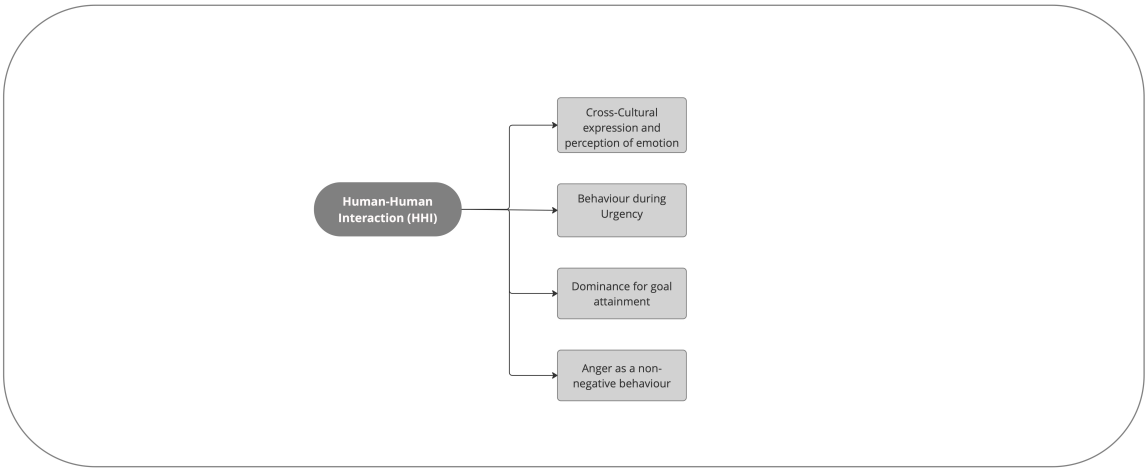


Figure 2.2: Approach of Literature review for HHI

2.1 Healthcare Human-Robot Interaction

Robotic technology presents a promising solution that can operate independently or with minimal human intervention to address the impending scarcity of healthcare workers. Initially designed to undertake labour-intensive tasks that are dirty, dangerous, demeaning, degrading, or driving with high precision, robots are now being deployed in various healthcare applications, including surgical and non-surgical procedures, patient care, and hospital logistics [1] [3], [25], [26]. The first documented instance of robot-assisted surgery dates back to 1985 when a robotic arm was connected to a Computerized tomography (CT) scanner for a CT-guided brain tumour biopsy [25].

Subsequent advancements in robotic technology, particularly in artificial intelligence (AI) and computer vision, have revolutionized healthcare robotics [25]. The global healthcare robot market is projected to reach a value of $12.7 billion by 2025 [2]. Today's healthcare robots include surgical robots like the da Vinci system that perform precise surgical procedures with enhanced dexterity. Social robots such as Pepper and Nao by Softbank Robotics provide companionship and assistance in caregiving settings [25], [27]. These robots aid with duties ranging from precision surgical procedures to the transportation of supplies and medications, enhancing operational efficiency and addressing the shortage of healthcare professionals.

Logistics robots, in particular, play an important role in supporting hospital staff. Robots like Moxi [28], [29] and ABB's YuMi [30] perform jobs like collecting supplies and sorting test tubes [30]. Aethon's TUG transports prescriptions and supplies

[1] https://www.agvnetwork.com/hospital-robots
[2] https://www.ahu.edu/blog/robotics-in-healthcare

without a robotic arm [3], and Swisslog Healthcare's Relay autonomously distributes medicines and lab samples, avoiding obstacles and lowering the chance of human error (Kyrarini 2021). Integrating these robots into hospitals improves service quality, staff productivity, work happiness, and cost savings, and allows healthcare workers to focus on patient care [25].

As robots become more prevalent in healthcare, effective human-robot interaction (HRI) strategies become crucial, including verbal and non-verbal cues, are crucial for robots to be perceived as valuable and acceptable in medical facilities [31], [32]. Social robots which are a type of service robots, in particular, must exhibit human-like sociability, express emotions, engage in conversations, and use natural communication cues to foster positive interactions [4]. It is because humans, tend to anthropomorphize objects, and in the case of robots, it increases with familiarity with the robot [33], [34].

As Breazeal et al. [35] said, social robots need *"to be social as we are"*. To achieve this, robots must adopt behaviours and communication styles that align with human social norms, fostering interactions that are perceived as friendly, helpful, caring, and trustworthy [36]. Conversely, rude, conceited, or hostile behaviours must be avoided to prevent negative perceptions and feelings among human users [36]. Given the diverse user profiles in healthcare settings, robots must be capable of safe interaction, behavioural interpretation, and transparent communication of their internal states to integrate seamlessly into human-centric environments [13].

To achieve this, researchers and industry professionals are exploring the field of HRI) [11]. HRI focuses on designing, understanding, and evaluating intuitive interactions between humans and robots [11]. This can occur through teleportation, full automation, or implicit communication [20]. HRI brings together multiple disciplines, including robotics, engineering, computer science, human-computer interaction, cognitive science, and psychology [37].

Social HRI, in particular, focuses on creating machines that possess social capabilities and can interact with people naturally and intuitively [13]. To achieve intuitive social HRI, robots should be designed to evoke social cognition in humans as intentional agents, influenced by HHI [38], [39]. This involves utilizing a range of the most important modalities in HRI: visual, auditory, and tactile, often presented using visual displays, gestures, natural language (text and spoken language), audio, physical interaction, and haptics [11].

Factors such as the proximity of humans and robots, the robot's voice and the robot's morphology [13] could also play a role in communication strategies. Robot voices [40] can be categorized into speech (natural language) or spoken language

[3] `https://aethon.com/mobile-robots-for-healthcare/`

[4] `https://robohub.org/understanding-social-robotics/`

similar to human voices, and robot sounds, which are non-verbal sounds like semantic-free utterances (SFUs discussed in Section 2.3.1) and other noises like motor sounds robots emit [11], [37], [41].

Another important aspect to recognize is that the relationship between society and technology is complex and mutually influential [42], and both elements continuously shape each other. Hence, it is important to study each other's influence to design effective and intuitive strategies for integration of robots within healthcare settings.

In the field of HRI research, users' perceptions of robots are influenced by various factors such as cultural background, age, education, and exposure to robots through literature, entertainment media, or personal encounters [6], [43], [44]. Challenges related to the social acceptance of artificial agents and differing perceptions based on cultural contexts have been identified by researchers in HRI [43]. Globalization also plays a significant role in shaping users' perceptions [6]. This is discussed in Section 2.2 in detail. Therefore, effective communication strategies are essential at both social and individual levels to enhance the social interaction experience between humans and robots.

In conclusion, as Human-Robot Interaction (HRI) is an interdisciplinary field, it is crucial to carefully design, develop, or select communication strategies that take context, cultural backgrounds, and prior experiences into consideration to ensure global social acceptance, adoption, and successful user interactions. Robots present promising solutions to healthcare challenges like labour shortages, and operational inefficiencies, along with the safety of healthcare workers, but their success depends on more than just technological capabilities. As healthcare robots evolve, effective HRI strategies by paying attention to robot's social behaviour, particularly in culturally diverse environments, is essential for robots to be accepted worldwide, trusted, and capable of providing safe, effective care.

2.2 Culture Definition

Culture is a complex and multifaceted concept. In the field of social science, it is defined as:

> ”Culture is a mould in which we all are cast, and it controls our daily lives in many unsuspected ways.” - Hall [45]

In the field of design, culture is defined as [46]

> ”Culture is [the] system of shared beliefs, values, customs, behaviours and artefacts that the members of society use to cope with their world and

with one another, and that are transmitted from generation to generation through learning".- Krober and Kluckhohn [47]

Krober and Kluckhohn [47] identified it's 164 definitions as mentioned in a paper by Van Boeijen [46] which emphasizes it's diverse interpretations.

Culture significantly impacts how individuals perceive and interact with technology, influenced by factors such as nationality, religion, race, and socioeconomic status [6]. An individual's mental model, encompassing cognition patterns, emotions, and behaviour, is moulded during the formative years by early childhood experiences, family dynamics, social environment, and broader societal contexts [48]. Hofstede [49] describes culture as *"collective subconscious programming of the mind"*, influencing thoughts and actions based on past experiences [24], [48], [50]–[52].

Hofstede's key cultural dimensions, such as individualism and collectivism, affect our cognitive processes and psychological experiences [23], [24]. Other cultural dimensions, proposed by Hofstede, including power distance, uncertainty avoidance, and masculinity-femininity, also help in explaining these differences in emotional expression [9], [49], [53], [54]. These key cultural dimensions relevant to this study are explained below in brief:

Table 2.1: Hofstede's Key Cultural Dimensions Relevant to this Study

Dimension	**Description**
Individualism-Collectivism	This dimension assesses how individuals are integrated into groups. Individualistic cultures value personal independence and individual rights, whereas collectivist cultures emphasize social cohesion and common well-being.
Power Distance	This dimension measures acceptance of unequal power distribution. High power distance indicates acceptance of hierarchical structures, while low power distance suggests a preference for equality.
Uncertainty Avoidance	This dimension measures comfort with ambiguity. High uncertainty avoidance cultures prefer clear rules and stability, whereas low uncertainty avoidance cultures are more open to change and risk.
Masculinity-Femininity	This dimension examines gender role distribution. Masculine cultures value competitiveness and achievement, while feminine cultures emphasize relationships and quality of life.

Dutch society ranks high on individualism (Individualism Index = 100) and moderate on power distance (Power Distance Index = 38), emphasizing self-actualization, self-reliance, competition and equality among individuals. In contrast, Indian society ranks low on individualism (Individualism Index = 24) and high on power distance (Power Distance Index = 77), reflecting strong family ties, collective decision-making, respect for hierarchy, and deference to authority [5].

Triandis's framework [55] emphasizes the influence of an individual's age and gender on their perception. Research on Dutch society suggests that individualism tends to increase with age, although, in collectivism, it tends to decrease [56]. As individuals transition from family-oriented to educational and professional environments, shifts in attitudes occur. Additionally, gender differences play a role, with females catching up with males in understanding societal emphases on individualism and collectivism by about 22 years of age, similar to males [56].

Moreover, urbanization and globalization, in addition to national culture, can influence individuals' perspectives [57], particularly evident in third-world countries or countries affected by colonialization, such as India [58]. Given that culture influences various aspects of human behaviour, norms, and perceptions of the world [48], understanding these cultural dimensions becomes crucial for designing effective HRI strategies and facilitating global acceptance of robots in public domains. The selection of cultural dimensions to focus on completely depends on the specific research [59].

2.2.1 Cultural in Communication and Interpretation of Emotions

The debate regarding whether emotion perception is universally shared or culturally distinct has been a long-standing topic of discussion [60]. Studies have shown that cultural variations can influence nonverbal behaviours across different societies [8], [61]. Although Darwin proposed that emotions and their expressions are universal [62], subsequent research indicates that cultural differences shape how emotions are expressed and perceived [60].

Ekman [63] and Izard [64] from their research also support that basic emotions, including anger, are universally recognizable through facial expressions [60]. Although emotions are universally experienced, their expression and social interpretation vary due to differing cultural frameworks [65]. Subtle cultural differences affect how emotions are perceived and processed [60].

Cultural contexts and their established norms [66] influence communication styles, personalities, cognitive frameworks, and motivations [5], [67]. These differences affect judgments and the perception of emotional intensity [68]. This may be due to

[5] https://www.hofstede-insights.com/country-comparison-tool?countries=india%2Cnetherlands

variations in acoustic properties like pitch, loudness, and length of speech components that reflect the speaker's emotional state, making it easier to identify basic emotions across cultures using vocal cues [65].

Research indicates cultural variations in expressing negative emotions like anger [65]. Cultural dimensions such as masculinity and femininity, individualism and collectivism predict emotional expression [9], [69]. Individualistic societies like the Netherlands favour direct verbal expression, however, collectivist cultures like India prioritize indirect communication and nonverbal cues [5], [69]–[71].

A study by Hareli [57] found that angry and sad expressions were perceived as more intense in Germany and Israel than in Greece and the US. Zhang and Pell [65], referencing Scherer et al. [72], found that participants from Europe, America, and Asia could accurately identify emotions in German (anger, sadness, fear, joy) 66% of the time, significantly higher than chance. Negative emotions, especially sadness and anger, were recognized more accurately than joy or happiness [73]. Zhang and Pell also found that Canadian and Chinese participants were more accurate at identifying emotions in their native languages and perceived negative emotions like anger as more intense, indicating similar inferential rules across cultures. Some research suggests that anger is one of the negative emotions with high recognition rates for vocal emotional recognition [74] as well as non-verbal vocal signals across different cultures [75].

Recent research also indicates that the impacts of urbanization and globalization on individuals, which lead to exposure to diverse cultures, can also influence the expression and perception of emotions [6], [57]. This shift in perception and preferences can be attributed to exposure to diverse cultural communities, leading to changes in how individuals perceive and prefer certain things [10]. This could be because cultural identity changes over time.

In conclusion, understanding the influence of cultural differences on communication and interpretation of emotions is crucial in diverse social contexts, especially when designing for social interaction. Embracing cross-cultural perspectives can lead to more effective strategies for designing human-robot interaction systems sensitive to diverse cultural norms and preferences, enhancing usability in public spaces.

2.2.2 Culture in Human-Robot Interaction

In the field of HRI, culture is often defined by *Nationality* [8]. In HRI research, scholars frequently reference Hofstede's cultural dimensions, particularly individualism and collectivism [8], [43], as well as concepts of femininity and masculinity, and Hall's theory of low-high context [24], [49] to delve into cultural influences to understand individuals' responses and perceptions from diverse culture worldwide.

Hall theory [24] states that High-context cultures communicate indirectly and prioritize harmony, making them better at establishing and maintaining connections through cohesiveness within society. Low-context cultures, on the other hand, are more direct and problem-focused, which leads to improved efficiency in accomplishing individualistic goals by emphasising clarity and directness in communication, according to Hall.

The emphasis on cultural variables in robot design and development is crucial [68], [76]–[78]. Individualistic cultures emphasize personal significance, although collectivist cultures value group harmony, impacting how robot communication is perceived [8]. For example, cultural differences between East and West in the use and design of virtual avatars and robots have been extensively studied. Attitudes towards robots also differ across regions, with robots perceived as enemies in Europe, servants in America, colleagues in China, and friends in Japan [79].

Variations in human speech patterns have also been observed in interactions with virtual avatars, with Korean speakers exhibiting more negative behaviour influenced by collectivist communication styles compared to English speakers [77]. These variations in perception of speech patterns are due to culture. Gonzalez's study [19] explored the functionality of gibberish speech (a type of Semantic-Free Utterances, as defined in Section 2.3.2) through the "Talk to Kotaro" platform, analyzing responses from individuals of diverse cultural backgrounds. Despite limitations in sample size ($n = 13$), the study emphasized the importance of considering cultural differences in HRI.

On the other hand, the evolution of technology has a reciprocal (bidirectional) relationship with culture, as both influence each other [42] making it essential to explore cross-cultural perceptions when designing and developing any technology (robotic in this case). Further research is needed to fully understand the significance of cultural factors in HRI within an increasingly globalized society, as the outcomes have been inconclusive so far [43], [66]. Therefore, understanding cross-cultural perceptions towards robots is essential in designing culturally acceptable robots, especially for public spaces [5] to foster more inclusive and effective HRI experiences.

As social signals are also heavily context-dependent and culturally driven, the robots deployed in human environments must have the capability of adapting to cultural differences and learning the appropriate social and moral norms [4]. So, given the diverse perceptions and attitudes towards robots across different cultures, it is crucial to study the perception of robot communication for potential employment in public settings such as hospitals, within a cross-cultural context. Therefore, the present study aims to explore cultural factors by analyzing participants' perceptions of two distinct cultural groups, i.e., individualistic and collectivistic.

2.3 Semantic-free Utterances

Sound, in particular, plays a crucial role in shaping the impact and user experience during social interactions by conveying the robot's internal state, eliciting emotions, and influencing perception, even at a cultural level [80]–[83].

Although natural language is the predominant method of communication used in robots [14], designing HRI systems that can effectively communicate with users from diverse cultural backgrounds using natural language poses several challenges. [13]. The current state of the art in natural language technology is still not there yet from machines holding open-ended conversations on par with young children, let alone adults [13].

Natural language processing (NLP) involves a "serial pipeline method" that includes tasks such as tokenization, part-of-speech tagging, named entity recognition, syntactic parsing, semantic analysis, and more [13]. The process is linear, leaving very little room for error. A mistake in any stage can propagate quickly, leading to a breakdown in the Natural Language Interfaces (NLI) and interfering with effective communication in robotics systems [13].

Research suggests that people from different cultural backgrounds tend to prefer communication from a robot with neutral expressions rather than culturally aligned ones [66]. In public spaces where people from diverse cultures interact, designing HRI with the natural languages of over a thousand cultures could be challenging. This highlights the importance of addressing culturally neutral expressions or expressions that cater to the diversity in HRI. In this regard, semantic-free utterances (SFUs) offer a promising solution for communication in multicultural contexts, as well as with pre- or non-lingual individuals [13].

Social human-robot Interaction (HRI) draws inspiration from established fields like animation. Characters like R2D2 from Star Wars and WALL-E from Disney-Pixar show that robots can effectively communicate without natural language. These examples demonstrate how alternative methods of communication like SFUs can engage and entertain audiences, inspiring the future of social robots in various real-world contexts.

2.3.1 Defining Semantic-Free Utterances:

Semantic-free Utterances (SFUs) are a type of auditory communication and vocalizations that convey emotions and intentions without semantic content or relying on specific words from any language [13]. *"SFUs comprise everything which is not the* *focus of semantics, syntax and morphology in natural language and which does not include instrumental and natural sounds in acoustics and music" [13].* SFUs are not limited to a specific spoken dialect and can effectively convey emotions and intentions without relying on complex linguistic structures [13].

SFUs include various aspects of auditory communication such as speech, language, vocal factors, non-vocal sounds and acoustics [84]. Although language pertains to natural language processing within computational linguistics, speech refers to spoken language analyzed by speech processing technology [13]. Vocal factors such as pitch, loudness, rate, pitch contour, and voice quality contribute to the expressive qualities of SFUs [13]. Acoustics, on the other hand, focuses on non-vocal sounds within the audible frequency range, including instrumental and natural sounds and computer-generated sound wave effects [13].

2.3.2 Classification of Semantic-Free Utterances

Semantic-free utterances are classified into four general types (figure2.3): Gibberish Speech (GS), Non-Linguistic Utterances (NLUs), Musical utterances (MUs), and Para-linguistic utterances (PUs) [13].

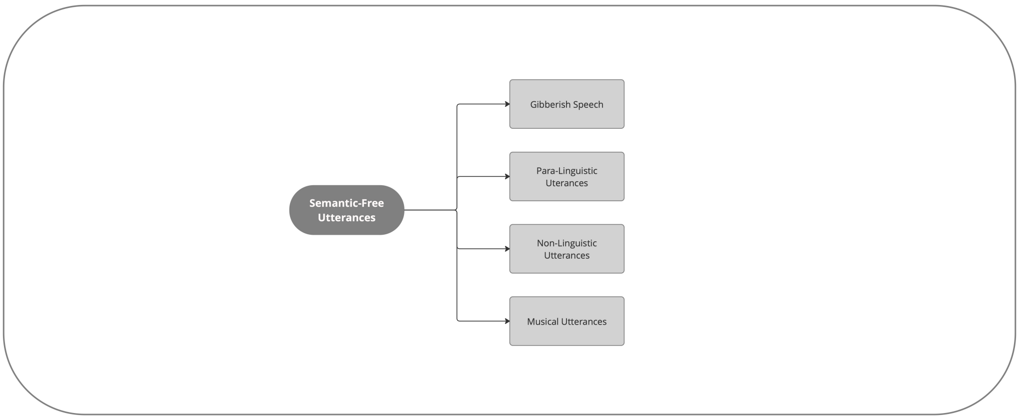


Figure 2.3: Classification of Semantic-Free Utterances

NLUs and MUs are non-vocal sounds, whereas GS and PUs are vocalizations of human speech [13]. In this thesis, the focus is on GS and NLUs, which are discussed in detail below:

SFUs have demonstrated their effectiveness in a variety of media and commercial applications. They have been featured in popular films such as Star Wars, Wall-E, and Minions [13]. They have also been incorporated into commercial robots like Anki's Cozmo and Vector, which were inspired by Wall-E [88]. These robots, designed for children aged 8 to 14, illustrate the potential of SFUs to enhance expressive displays and advance research in human-robot interaction. Despite lacking

Table 2.2: Types of Semantic-Free Utterances and Their Descriptions

Type	**Description**
Gibberish Speech (GS)	Gibberish speech refers to vocalizations that lack semantic content but effectively communicate emotions and affects [13], [85]. It can resemble speech and may involve systematically encrypted language or speech that is semantically nonsensical but grammatically correct [13]. Two main strategies at the speech level are cue/content masking and cue/content manipulation through synthesis. At the text level, strategies include Chomsky's approach and the Jabberwocky sentence approach, which destroy semantics at the sentence and word level, respectively [13]. Yilmazyildiz [85] demonstrated that GS could convey emotions effectively, with 81% recognition accuracy in HRI [13].
Non-linguistic Utterances (NLUs)	NLUs refer to non-speech sounds that do not resemble speech, such as beeps and squeaks [86]. They are used to convey emotions and information with varying accuracy and are often cheaper and easier to implement than language-based systems [14]. Approaches to NLUs include sonification, auditory icons, and earcons. Sonification converts data to sound, auditory icons mirror real-world sounds, and earcons are synthetic patterns representing specific events [13]. Liu [87] found that NLUs, like GS, can convey basic emotional expressions with varying accuracy, highlighting their potential in HRI.
Musical Utterances (MUs)	MUs are non-speech sounds derived from music theory, adjusting parameters like tempo, pitch, melody, and rhythm [13]. They can convey emotions, intentions, or statuses and are useful in noisy environments where spoken communication is difficult [13].
Paralinguistic Utterances (PUs)	PUs are vocal or nonverbal signals that extend beyond verbal messages, including pitch, rhythm, and vocal quality, as well as nonverbal clues such as gestures and facial expressions [13]. PUs enhance emotional expression and user engagement in HRI, improving the overall interaction experience. Despite their growing popularity in speech processing, their use in HRI studies remains limited [13].

natural language, these films and robots have been widely understood and appreciated across diverse cultures, underscoring the effectiveness of SFUs as a form of communication [13].

In academic research, SFUs are a focus of study, particularly in the context of child-human-robot interaction (Child Human-Robot Interaction (cHRI)) [89]. For instance, studies with the Kismet Robot [90] which uses babbling vocalizations similar to GS, indicates that individuals could effortlessly participate in multi-modal proto-interactions with the robot without relying on natural language interaction [13].

Further research indicates that human listeners had a low rate of detecting certain emotions, such as fear in gibberish-like speech in HRI [35]. This suggests that the uncertainty experienced in natural language regarding emotions also applies to gibberish-like speech in HRI. Findings from this study indicate that humans perceive gibberish-like speech as a natural language, indicating its potential use for communication in certain cases of human-robot interaction, particularly in the field of (cHRI).

These findings point to the promising role of SFUs in facilitating communication in HRI. This thesis aims to explore the potential of SFUs for interacting with a hospital logistic robot.

2.3.3 Benefits of using Semantic-Free Utterances in HRI

Natural language interactions often raise expectations regarding the social and cognitive capabilities of conversational partners, potentially resulting in disappointment in certain situations. To avoid this, utilizing SFUs in communication has been recommended as a way to manage expectations when interacting with artificial systems, as opposed to natural language interactions [13], [81], [91], [92]. This is because in HRI, robots create their social environment with their dynamics [93], and using natural language and imitating human voices may not always be necessary [34]. In this context, HRI designers can subtly regulate user expectations of robotic systems by producing robots that do not engage in open-ended Natural Language Interaction (NLI) but still respond to external stimuli when using SFUs [13].

Research has shown that individuals tend to have more trust in robots that utilize synthetic or mechanical-sounding voices more than those with human-like voices [40]. This preference may stem from the perception that robots employing artificial language or SFUs instead of natural language are less intelligent [13], thus reducing fear about them. By utilizing SFUs, robots can maintain a machine-like identity, ensuring their functional efficiency without causing discomfort to humans who might find overly human-like robots unsettling (uncanny valley, [6] or Automatono-

[6] https://en.wikipedia.org/wiki/Uncanny_valley

phobia [7]). These findings underscore the potential benefits of SFUs in promoting positive interactions between humans and robots.

Notably, SFUs, particularly NLUs, do not require high cognitive abilities for user comprehension [94], unlike NLP interfaces, which may demand advanced cognitive abilities for interpreting textual and spoken utterances. It is because SFUs contain less semantic information to process, reducing the cognitive load on users when processing the robot's speech [13]. This aids in maintaining fluidity in vocal conversations, a crucial aspect of HRI [13].

An example akin to SFUs is the development of artificial languages like Robot Interaction Language (ROILA), which optimizes speech recognition performance by requiring minimal user effort to learn, however achieving high accuracy on the robot side [13]. By incorporating SFUs in robots, the cognitive load on humans in HRI can be reduced, leading to smoother conversational interactions with a lower risk of breakdown compared to NLP interfaces [13].

In conclusion, although incorporating natural language interfaces in robots requires sophisticated speech recognition, language understanding, and cognitive modelling [20], using SFUs can be a more accessible and adaptable choice, particularly in public settings with diverse cultural and age demographics, like hospitals, and when conveying unintentional communication. Additionally, SFUs can be useful for communicating alerts or warnings, as non-speech expressions (auditory icons specifically) have been used in this context for ages [20]. This thesis will explore this specific aspect in the following sections.

2.4 Urgency

In this section, we will explore urgency and its effective communication that involves using various expressions and behaviours, such as anger and dominance, to be incorporated into the robot, ensuring that these signals do not come across as rude.

2.4.1 Defining Urgency

The concept and interpretation of urgency are multifaceted and can exhibit variations based on the specific context. According to the Cambridge Dictionary, *Urgency* is described as *the quality of being very important and needing attention immediately"*[8], indicating a pressing need that demands prompt and efficient action[9]. Consequently, urgency is commonly linked with concepts of rush or hurry,

[7] `https://www.piernetwork.org/automatonophobia.html`

[8] `https://dictionary.cambridge.org/dictionary/englis/urgency`

[9] `https://www.reallygoodinnovation.com/glossaries/urgency`

particularly with subjects concerning movement [10].

In the context of this research, urgency is defined as the effective conveyance of a sense of haste by a robot engaged in delivery tasks. Traditionally, auditory warning signals have been employed to communicate urgency due to their distinct advantages over visual and tactile alerts [95]. This choice is informed by the fact that auditory cues are omnidirectional, cannot be involuntarily ignored and generally exhibit superior performance in eliciting quicker response times, whereas visual warnings necessitate direct sight for effectiveness [95].

2.4.2 Expressing Urgency

Various sounds can communicate different levels of urgency, with perception influenced by the specific situation or task at hand [96]. Urgent warning sounds should possess characteristics such as alertness, persistence, and attention-grabbing ability to facilitate swift recognition of a warning situation [97]. Common warning sounds like horns, bells, buzzers, and sirens are used to capture attention and emphasize the need for immediate action [98]. Non-linguistic sounds, such as auditory icons and earcons, serve as effective tools for communicating urgency in human-machine interfaces. These signals provide discreet and universally understandable warnings that transcend language barriers and are culturally relevant [99]. However, there have been inconsistencies in participants' perceptions of urgency when exposed to auditory icons, particularly in in-congruent scenarios [99].

In the field of HRI, the effective communication of urgency can be influenced by various factors, including voice characteristics and cultural nuances [100]. Ensuring appropriate urgency communication is essential, especially in time-sensitive tasks where robots may need to take precedence over humans and assert dominance [101] like delivering bio-samples in a hospital.

In a research study examining the communication of urgency in HRI, a thorough sound design approach was utilized to create sounds known as SFUs that effectively convey urgency by robots [2]. The study specifically focuses on how robots can communicate urgency during unintentional interactions, which is crucial for users to understand the need to give space to the robot without perceiving the robot's behaviour or sounds as impolite or alarming. These sounds are designed to resemble auditory warnings, which could be mistaken for alarms or emergency alerts in hospital settings, potentially causing confusion among individuals.

In conclusion, effective urgency communication is essential in critical tasks such as healthcare, where clear communication can be a matter of life and death. With the integration of robots in such environments, it is imperative to explore communication

[10] https://thesaurus.plus/related/hurry/urgency#sentences

strategies that differentiate urgency signals from other environmental sounds, such as ambulance sirens or hospital background noise. The use of SFUs, as discussed in Section 2.3, holds promise in addressing this challenge.

2.4.3 Expressions and Behaviours for Expressing Urgency

Dominance

One key behaviour that has been identified as expressing urgency is Dominance. This term generally refers to taking control or influencing others to accomplish tasks, often by showing aggression to gain priority [11][12]. It is a situation-dependent behaviour in which one individual seeks to influence, regulate, or control the behaviour of another [101] by exercising power and influencing ideas, feelings, or actions [102], [103].

Dominance is an implicit yet crucial strategy in social interactions, conflict resolution, and establishing hierarchies [101], [104]. Research indicates that those with higher levels of dominance often have greater social influence and control over conversations [103], [105].

Dominance in technology has been a crucial factor influencing psychological responses to computers [103], [106]. Prior research in acHRI has explored non-verbal indicators such as eye contact, proximity, body postures, gestures, and motion to convey dominance [104]. However, sound has not been thoroughly explored. For instance, a robot's motion can be perceived as dominant when it continues its task despite potential interference, prompting humans to prioritize the robot [101].

A study demonstrated that hospital staff perceived a sense of disrespect when robots were prioritized in hallways through dominant motion patterns [107]. It is important to consider the impact of dominant interactions on human perception, as they can adversely affect the perception of robots in HRI and their social acceptability.

In conclusion, a robot's dominant behaviour can be further investigated to convey a sense of urgency through various modalities. This allows for exploring vocal modalities beyond just movement or behaviour usually studied using visual cues. This aspect forms the focus of the thesis in determining a connection between urgency and dominance.

[11] https://www.dictionary.com/browse/dominance,
[12] https://dictionary.cambridge.org/dictionary/english/dominance

Anger

Anger is often viewed as a negative emotion driving aggressive or hostile behaviour. However, research indicates that anger can also lead to beneficial and prosocial behaviours, fostering collaboration and moral behaviours [108].

Functional accounts of emotion theories suggest that emotions serve a purpose in responding to challenges [109]. Different emotions can be viewed as tools to solve specific problems [109]. For instance, similar to dominance (as discussed in the previous section 2.4.3), anger expression can influence the behaviour of others, leading them to back down or support the angry individual, making it useful in achieving goals [109]. Anger is often associated with a state of readiness for action, prompting others to remove themselves as obstacles [109]. In situations such as negotiations, displaying facial and verbal expressions of anger can influence cooperation and lead to favourable results [108], [109].

Research indicates a relation between anger and dominance in HHI. Anger is often seen as a marker of dominance, suggesting that anger may have evolved to enhance social negotiation, which could form dominant hierarchies [110]. As a result, dominant individuals are often perceived as angrier [111], which considerably impacts human social relationships.

Findings from multiple studies suggest that individuals from individualistic cultures experience anger more frequently than those from a collectivistic culture [112]. This could be because the differential focus of self-hood in individualistic and collectivistic cultures influences cultural standards of right and wrong and, therefore, emotional responses of approval and disapproval.

In conclusion, a relationship exists between anger and dominance, but most studies have primarily focused on visual cues of dominance in HHI [110] and even HRI. However, not a lot of evidence was found of this association that incorporates vocal cues or non-verbal vocal communication. Also, there is a lack of literature when it comes to finding a relation to anger and culture in HRI but there is in HHI when it comes to cross-cultural findings dealing with anger.

Rudeness

As discussed in subsection 2.1, negative perceptions such as rudeness, conceit, or hostility must be avoided to ensure that robots are socially accepted [36]. This is particularly important when evaluating the utilization of vocal interactions (SFUs in our case) that convey anger, have the potential to convey dominance, and can consequently be perceived as impolite or rude.

There are instances HHI in management psychology where anger-related behaviours are considered disrespectful or Rude. Cultural factors, such as honour cultures (collectivist culture [113]), may link anger to rude behaviour where anger is frequently expressed in response to rudeness or mistreatment [114]. Kim [115] in its study, explores how cultural differences influence individuals' reactions to rudeness. In collectivistic cultures like Korea, subordinates might react differently to supervisor rudeness compared to individuals in individualistic cultures like the USA. Despite this, the emotional response to anger and its association with perceived rudeness remain universal across cultures. However, the manner in which anger is exhibited may differ.

Given these findings, it is essential to investigate how the design of robot behaviours affects human perception negatively, particularly when using anger or dominant expressions in human-robot interaction. For example, a study involving a quadruped robot used for transporting goods in a hospital corridor that uses non-verbal behaviours found that participants perceived the submissive robot as safer than the dominant one, and some even perceived the behaviour to be rude [116].

This implies that perceived dominance and its potential accompanying behaviours, such as anger, could have a negative impact on the social acceptability of robots. Identifying these dynamics is also important for designing robots to be employed in social settings without evoking negative reactions, such as rudeness or hostility, from potential users. Although this study focuses on the same context of hospital logistics delivery robots, it utilizes a different morphology of the robot, and its mode of interaction is different with no audio interaction involved.

Given that anger is both a response to and a component of perceived rudeness across cultures, as highlighted in previous literature, it is imperative to consider how sounds associated with anger and/or dominance might influence the overall perception of robot behaviour. Therefore, it is essential to evaluate rudeness alongside anger and dominance to ensure that the robot design is socially acceptable across different cultures.

In conclusion, this study aims to address the gaps identified in the literature review in the use of SFUs, particularly in conveying urgency by hospital logistics robots, as well as to understand SFUs perception in diverse public settings. The study will draw conclusions based on Hofstede's cultural dimensions and insights from HHI research. This study aims to bridge these gaps and contribute to the design and development of effective communication strategies for HRI in multicultural public spaces, particularly for hospitals. Through exploration of SFUs that express anger with the potential of expressing dominance, this study seeks to enhance our understanding of how these cues can efficiently convey urgency in hospital contexts. The rudeness of these cues will also be examined to ensure their social acceptability. Details on the sources of SFUs, Hypotheses, and Methodology will be presented in the subsequent Sections and Chapters of this study.

2.5 Research Objective

To investigate the perception of urgency in hospital delivery robots across different cultural backgrounds (Indian and Dutch) and SFUs from two corpora (EMOGIB and BEST), designed research questions followed by hypotheses to statistically test them.

(RQ:) What is the effect of the type of Dominant Semantic-Free Utterances from two different corpora (BEST vs. EMOGIB) on individuals from two distinct cultural backgrounds (Indian vs. Dutch) when these utterances are utilized by a hospital delivery robot? This main research question also leads to the formulation of the sub-research question: *SRQ_1 (Database Focused):* What is the effect of dominant SFUs from two different corpora (EMOGIB vs. BEST) on the perception of urgency in robot behaviour?

SRQ_2 (Culture Focused): What is the effect of dominant SFUs in robot behaviour on the perception of urgency across Dutch and Indian cultural backgrounds?

In this study, SFUs from two corpora, EMOGIB and BEST as utilized, as mentioned in the previous Section 3.1. These corpora comprise two distinct types of SFUs: Gibberish speech GS and Non-linguistic utterances NLUs, having distinct characteristics in their design and creation, as discussed in Section 2.3.2. The former resembles human speech, however, the latter has been used to convey urgency or emergency in auditory warnings (Section 2.3.2). As humans, tend to anthropomorphize objects, and in the case of robots, it increases with familiarity with the robot [33], [34], they might prefer EMOGIB over BEST. On the other hand, NLUs might evoke associations with emergency warnings or alerts, which are familiar contexts for individuals. Additionally, research has demonstrated that various non-verbal cues, including vocalizations, influence the perception of urgency [99].

Given the varying characteristics and emotional expressiveness of the semantic-free utterances from the BEST and EMOGIB corpora, it is reasonable to expect differences in their perception of urgency. Consequently, participants may perceive SFUs from these corpora differently in terms of urgency, depending on the specific characteristics of the sounds and their ability to evoke a sense of hurry as an intent communication. Therefore, to investigate this main effect of corpora on urgency perception, it is hypothesized that:

H1: In a hospital delivery robot, the urgency perception of semantic-free utterances from BEST and EMOGIB will differ, regardless of participants' cultural backgrounds.

Cultural differences significantly impact the interpretation and expression of emotions, including urgency cues [60]. For instance, collectivist cultures like India tend to prioritize indirect communication and nonverbal cues, whereas individualistic cultures like the Netherlands favour direct verbal expression [5]. Moreover, cultural dimensions such as power distance and individualism-collectivism influence emotional expression and perception [9], [69]. Given these cultural variations, participants from India and the Netherlands are likely to interpret semantic-free utterances differently in terms of urgency, reflecting their cultural norms and communication styles.

As discussed in the section 2.2.1, cultural variations play a pivotal role in shaping individuals' perception and evaluation of emotional intensity, resulting in divergent evaluations among different cultural groups [8], [68]. Sound interpretations may be affected by cultural differences [83]. Moreover, cross-cultural studies have underscored the disparities in emotional expression and perception, elucidating how cultural norms modulate the encoding and decoding of emotions as social cues, thereby manifesting variations across societies [57]. For instance, in collectivist societies like India, characterized by high-context communication, indirect and nonverbal forms of expression hold significance. Conversely, in individualistic cultures such as the Netherlands, direct and explicit verbal communication takes precedence over para-linguistic elements and nonverbal cues [5], [70].

Given the profound impact of cultural variations on emotional communication, it is imperative to acknowledge these differences when designing robots capable of effectively engaging with individuals from diverse cultural backgrounds [5], [8]. Based on these findings, it can be hypothesized that the perception of urgency in semantic-free utterances from a hospital delivery robot may differ among individuals from different cultural backgrounds. But there is no clear evidence on which country will perceive urgency more, so a non-directional hypothesis proposed is *H2: In a hospital delivery robot, the urgency perception of semantic-free utterances will differ amongst the participants of India and the Netherlands, regardless of Corpora.*

The interaction between database type and cultural background is expected to have a significant impact on urgency perception. Cultural differences influence individuals' interpretations of emotional cues, including those conveyed through nonverbal utterances [65]. Additionally, the characteristics of semantic-free utterances from different corpora may evoke varied responses across cultures. For example, participants from India (collectivistic culture) may perceive urgency differently when exposed to NLUs from the BEST compared to participants from the Netherlands, who may exhibit a different perception of the same (individualistic culture). There could also be no differences due to urbanization and globalization, especially among

Indians. Investigating the interaction effects of culture and corpora type is essential for designing effective communication strategies for hospital delivery robots to be operated in culturally diverse public settings.

H3: The interaction between the corpora (BEST vs. EMOGIB) and the cultural backgrounds (Indian vs. Dutch) will significantly affect the perception of urgency in a hospital delivery robot.

As outlined in Chapter 2, dominance can serve as a justified behavioural strategy in HRI when dealing with urgent tasks. This strategy can influence humans to prioritize the robot, thereby facilitating goal achievement when time is critical [101]. Although the potential of non-verbal cues such as body postures, gestures, and motion for conveying dominance has been explored, vocal interaction remains understudied [104].

Given the close association between vocal expressions of emotions (such as anger, fear, disgust, happiness, sadness, and surprise) and facial expressions [117], it's reasonable to extend this association to non-verbal auditory interactions like SFUs. Additionally, as discussed earlier, cultural factors and corpora can influence the perception of both urgency and potential dominance. Thus, these factors may affect the relationship between urgency and dominance.

Therefore, this study hypothesizes that in the context of a hospital delivery robot, the perception of dominance is positively correlated with the urgency conveyed by SFUs. This hypothesis aims to assess how cultural variations impact the perceived urgency and dominance in the robot's vocal cues, providing insights into optimizing HRI in urgent scenarios.

H4: In a hospital delivery robot, the perception of dominance is positively correlated with urgency conveyed by SFUs across different cultures.

This study will use a mixed-effects model (ART-ANOVA) to evaluate the main effects as well as the interaction effects (more details in Chapter 5.

Chapter 3

Databases of Semantic-Free Utterances

As seen in the previous section, this research aims to explore the capabilities of SFUs to be used in a hospital setting. As the hospital is a multicultural public space, this study is conducted with a sample from two distinct national cultures, India and the Netherlands. Typically, sound corpora contain expressions that convey basic emotions (anger, disgust, fear, happiness, sadness and surprise) [63]. Drawing inspiration from everyday HHI, this study sought to identify how doesSFUs that could effectively convey anger could also convey dominance and hence urgency for the service robot. By utilizing two types of SFUs: Non-Linguistic Utterances and Gibberish Speech from BEST and EMOGIB corpora (as mentioned in the next section 3.1) which have been investigated for conveying anger. This is performed in two parts: a pre-study and a main study. This chapter discusses the sources of stimulus, BEST and EMOGIB corpora for this study.

3.1 Corpora of Semantic-Free Utterances

3.1.1 Corpus 1: EMOGIB, the Gibberish Speech Corpus

Gibberish speech, despite the absence of comprehensible words, has proven to be an effective tool for expressing emotions, according to research by Yilmazyildiz et al. [13]. Even performing artists and cartoon animations, such as Teletubbies, have utilized gibberish speech to convey emotions without the use of actual words [16]. Refer to section 2.3.1 for more details.

The EMOGIB database is a corpus created for exploring GS in HRIn, comprising four distinct gibberish corpora. C1 and C3 utilize Dutch and English's whole consonant and vowel spaces. However, in Dutch and English, C2 and C4 use the entire vowel and voiceless consonant spaces. These corpora were recorded to convey affective child-robot interaction without needing meaningful words [16].

The EMOGIB database features an actress producing gibberish speech to convey six emotions: anger, disgust, fear, happiness, sadness, and surprise. Perceptual tests were conducted on all databases, which involved both adults and children, with recognition rates of up to 81% reported [16]. The EMOGIB database has significantly contributed to understanding the potential of gibberish speech in effectively conveying emotions. Therefore, further research on EMOGIB in different HRI contexts could provide valuable insights into its adaptability.

This thesis only uses sounds (GS) observed for conveying anger from this previous study [16].

3.1.2 Corpus 2: BEST, the Non-Linguistic Utterances corpus

As previously mentioned in Section 2.3.1, non-linguistic utterances have been identified as a category of SFUs (for more details, refer to Section 2.3.1) that rely heavily on human interpretation, which encompasses non-vocal sounds like beeps, squeaks, and whirls and is known to have relatively low semantic content [13].

The EU-funded Emote project (EMbOdied-perceptive Tutors for Empathy-Based Learning) [118] developed AI-powered systems to tutor children within a gaming environment. The systems recognize learners' emotions and respond with appropriate verbal and nonverbal expressions to foster empathetic interactions and positively impact engagement, affective reactions, and learning outcomes. The project also focused on the exchange of verbal and nonverbal cues and empathic affordances for effective communication.

Deliverable 3.2 of the Emote project focused on creating a validated corpus of non-verbal acoustical emblems (non-linguistic utterances) that establish a socio-emotional bond with the robot. These sounds effectively conveyed emotions such as happiness, fear, sadness, anger, disgust, and surprise and were evaluated using the Affect Grid. The project specifically targeted NAO robots that have limited facial expressions [118].

This thesis only uses sounds (NLUs) observed for conveying anger.

Chapter 4

Pre-Study

The study design in this research is structured into two parts: the pre-study and the main study. This chapter describes the pre-study phase, which involved the utilization of an online survey, and SFUs from the two corpora, BEST and EMOGIB.

4.1 Goal

The primary objective of the pre-study was twofold. Firstly, to investigate the potential relationship between selected Anger-expressing SFUs and their perceived Dominance among participants from India and the Netherlands.

Secondly, to streamline the sounds for the main-study from the pre-study, they were selected based on dominance ratings, reducing the number from 32 to 12. This included two sounds with high ratings, two near the median [119], and two with the lowest ratings from each corpus, as detailed in Chapter 5.

4.2 Study Design

The pre-study uses a within-subject design, where each participant was exposed to all conditions and the stimuli's order was randomized to mitigate potential order effects. The duration of the pre-study session spanned around 30 minutes and received ethical clearance from the ethics committee (EC-CIS) affiliated with the Faculty of Electrical Engineering, Mathematics, and Computer Science at the University of Twente.

The study was conducted using the Qualtrics online survey platform [1]. Participants at the start were introduced to the survey's purpose, including an explanation of SFUs and detailed instructions, as shown in Figure A.1. A visualization of a Harmony robot in a hospital lobby was provided to immerse participants in the context

[1] `https://www.qualtrics.com/`

where the auditory stimuli could potentially be used. It is a plain image with no person in the scene and just Harmony in the hospital space.

Participants provided anonymous online consent, as illustrated in Figure A.2 in Appendix A). Participants rated these audio stimuli for Dominance and Anger in part 1 of the pre-study which is discussed in Section 4.2.2 After rating the audio stimuli, participants completed part 2 of the study with a demographic questionnaire covering age, gender, national culture, and prior exposure to robots, discussed in the upcoming subsections. Detailed representations of the demographic survey can be found in Appendix A in Figure A.4, Figure A.5 and Figure A.6.

4.2.1 Stimuli

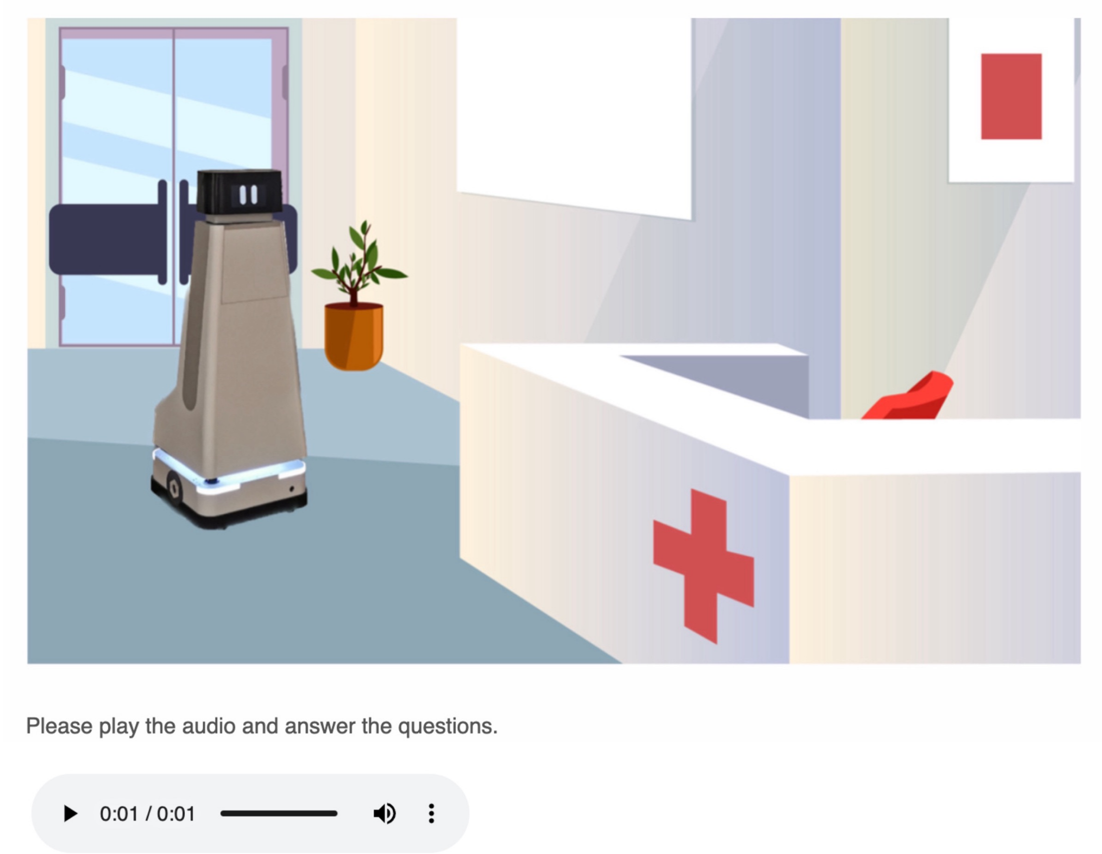


Figure 4.1: Stimuli Showed in Pre-Study on Qualtrics

Using Python's randomized function, a total of 32 sound stimuli (16 from EMOGIB and 16 from BEST) that conveyed anger were chosen at random from their respective corpus folders. Sixteen stimuli from the EMOGIB corpus were specifically selected from the folder of SFUs created using the English language to avoid potential language bias. These audio files were included in a Qualtrics survey along with an image of the Harmony Robot in a hospital lobby illustration, which served as a visual cue for participants (refer to figure 4.1). This simple illustration was created using Canva [2] and a stock-free image of the hospital lobby. The order of these SFUs was randomized to prevent any order effects.

4.2.2 Measurements

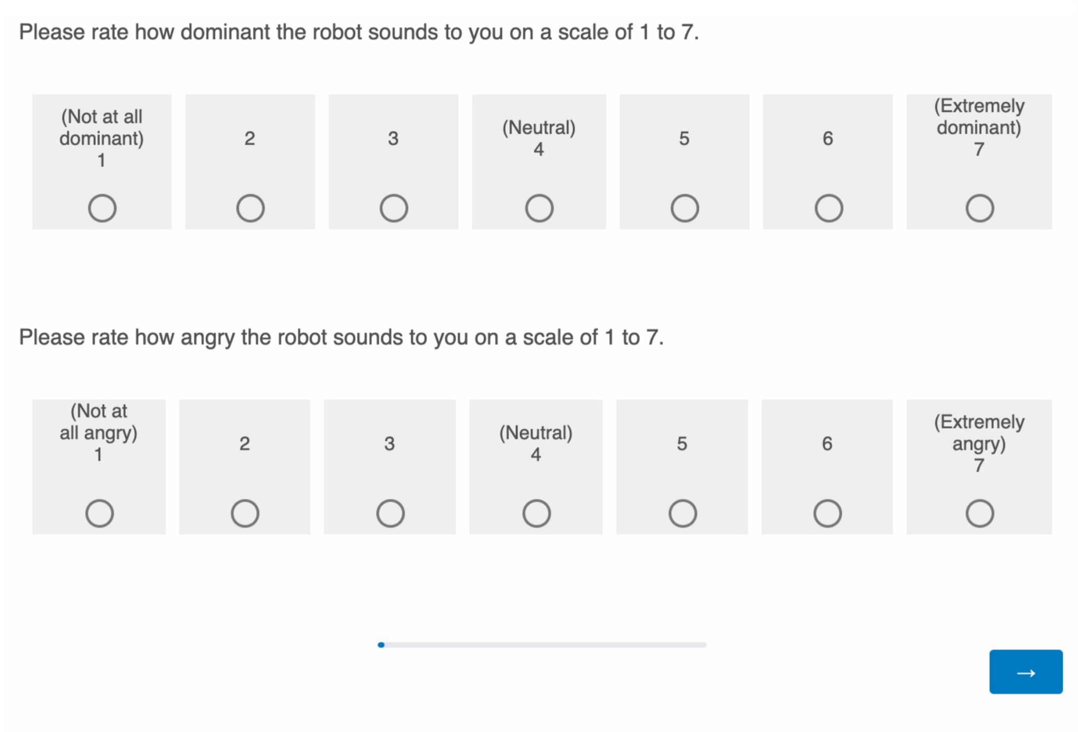


Figure 4.2: Scales used in Pre-Study on Qualtrics

The study's primary objective was to assess the dominance for each of these 32 SFUs. Participants were asked to rate the perception of the dominance of these SFUs for the robot on a Likert scale of 1 to 7, where 1 represented "Not at all dominant" and 7 represented "Extremely dominant". Participants were also asked to rate these sounds for the robot on the perception of conveying anger on a Likert scale of 1 to 7 where 1 represented "Not at all angry" to 7 represented "Extremely angry" as illustrated in Figure 4.2.

Participants were also asked for demographic information, such as which gender identity they identify with, their ages, the country in which they live, and whether their country of residence differed from their country of origin or their native country if "yes" and then which their country of origin and residence were, and if "no," only

[2] https://www.canva.com/en_gb/

their country of residence. They were asked to identify their national culture as indicated from Pre-study's Qualtrics in Appendix A in Figures A.4, A.5, and A.6. Participants were also questioned about their prior experience with the robots, using a Likert scale of 1 to 5, with 1 representing "Not at all Familiar" and 5 representing "Very Familiar".

4.2.3 Participants

The pre-study surveyed $n = 16$ participants (Indian = 8, Dutch = 8), recruited through a mix of purposive and random sampling through contacts and course study groups. Participation was voluntary and participants were required to be proficient in English.

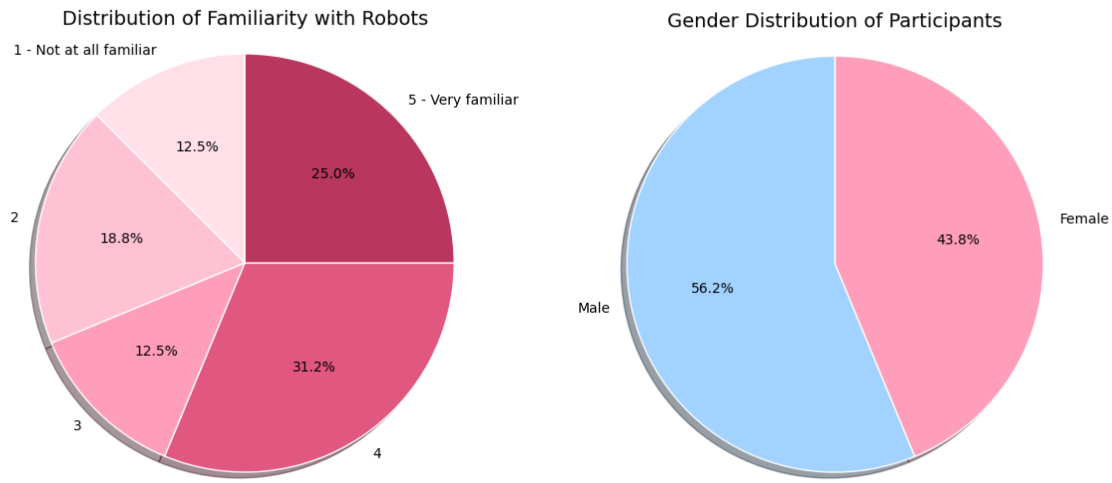


Figure 4.3: Pre-study Sample Infographic on Robot Familiarity and Gender Distribution

The sample (Figure 4.3 consisted of 56.2% ($n = 9$) identified as male and 43.8% ($n = 7$) identified as female, with a *mean* age of $M = 26.06$ years (*SD* $= 4.15$ years), ranging from 18 to 65 years.

Participants had varying degrees of exposure to robots as depicted in Figure 4.3. Among the participants, 12.5% reported having no familiarity rated 1 with robots however, 25% rated 5, very familiar. Overall, the majority of participants demonstrated a certain degree of familiarity with robots, as evidenced by the overall distribution of ratings in Figure 4.3.

4.3 Analysis and Results

A total of 17 responses were received, with one excluded due to lack of consent as the study was ended upon not giving consent. Data from 16 responses were analyzed using Python libraries pandas, numpy, scipy and matplotlib.

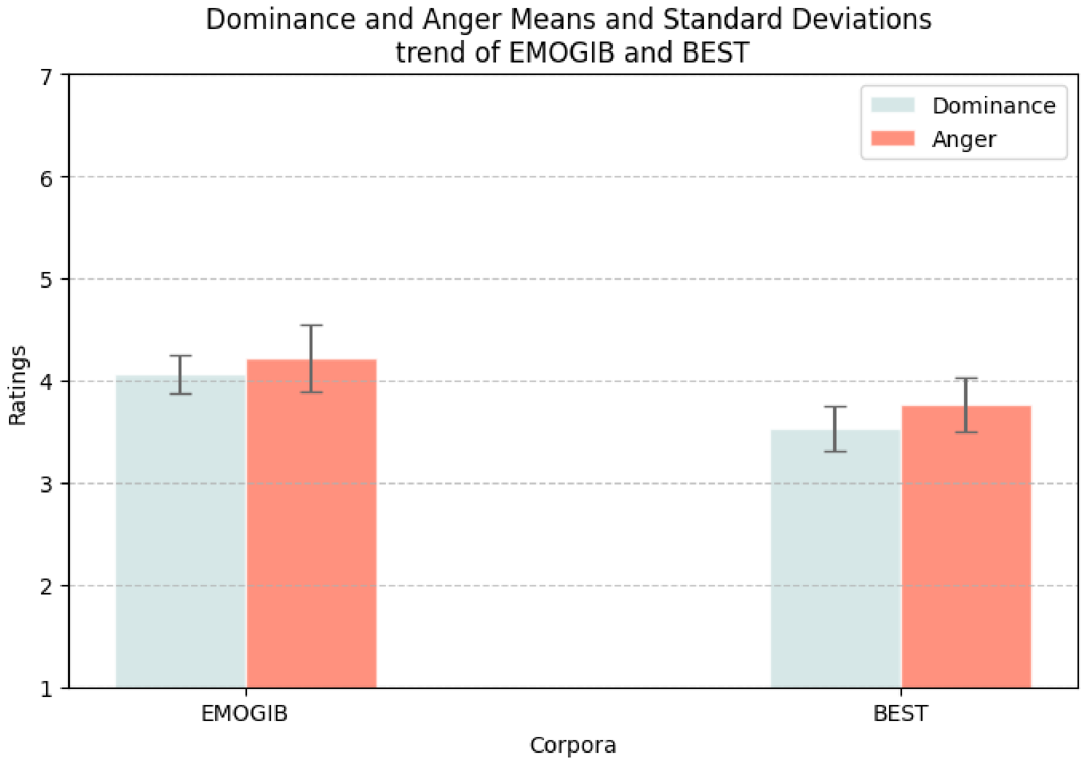


Figure 4.4: Overall ratings of EMOGIB and BEST for Dominance and Anger

The bar graph in Figure 4.4 represents the mean and standard deviations of dominance and anger across two corpora, EMOGIB and BEST. The EMOGIB corpus had higher mean ratings for both dominance ($M = 4.06$, $SD = 0.19$) and anger ($M = 4.22$, $SD = 0.33$) compared to the BEST corpus for dominance($M = 3.52$, $SD = 0.22$) and anger ($M = 3.77$, $SD = 0.27$). These higher mean ratings in the EMOGIB corpus suggest more intense perceptions of dominance and anger. Variability was low across both corpora, with slightly higher variability in anger ratings for EMOGIB, indicating slightly more varied perceptions of anger intensity. These results highlight a stronger emotional perception of the EMOGIB corpus, particularly regarding anger.

4.3.1 Relation between Dominance and Anger

For investigating the relationship between dominance and anger, the averages of ($n = 16$) participants' responses, were calculated separately for Anger and Dominance and its correlation was determined.

The results of this pre-study analysis revealed a strong positive correlation between Anger and Dominance perception for each sound, as indicated by a Spear-

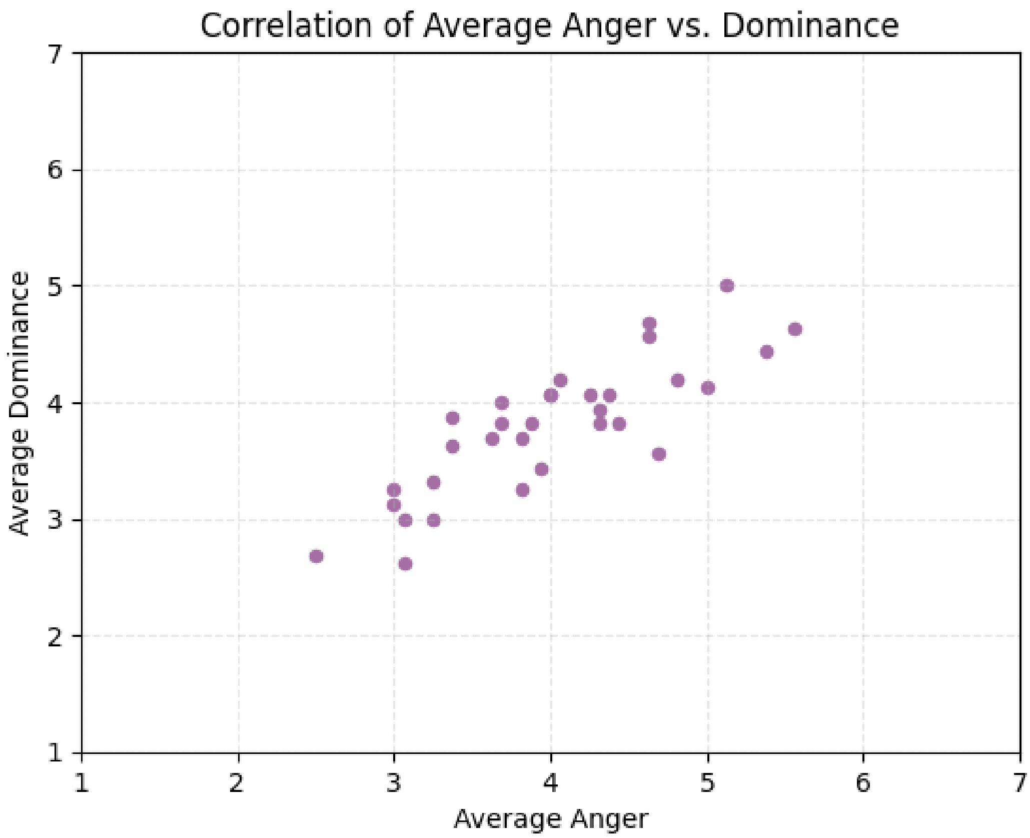


Figure 4.5: Correlation between anger and dominance

man's correlation coefficient of 0.829 ($p = 4.73 \times 10^{-9}$) as illustrated in Figure 4.5. This finding suggests that higher levels of reported Anger are associated with increased perceptions of Dominance. It is concluded that a positive correlation indeed exists between Anger and Dominance for the SFUs investigated from both corpora.

4.3.2 Sound Selection for Main-Study

Six sounds from each corpus were selected for Main-Study Chapter 5 based on their Dominance ratings as depicted from Figure 4.6 and Figure 4.7). These figures illustrate the mean and standard deviations of Dominance and Anger ratings for SFUs from the EMOGIB and BEST corpora, highlighting variability in the standard deviations.

From the BEST corpus (Figure 4.6), sounds were selected based on their Dominance ratings. The top two sounds rated highest for Dominance were Sound_3 ($M = 4.187, SD = 1.470$) and Sound_15 ($M = 4.125, SD = 1.892$). Sound_15 also exhibited higher levels of anger but showed greater variability in participant ratings ($n = 16$). For the medium-rated category, sounds close to a median of 3.500 were chosen: Sound_23 ($M = 3.562, SD = 1.459$) and Sound_26 ($M = 3.437, SD =$

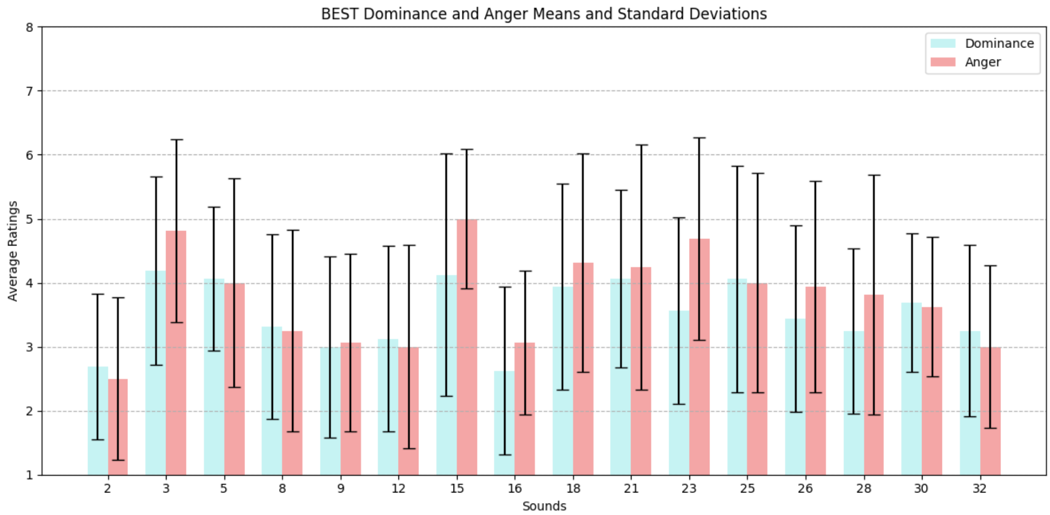


Figure 4.6: Dominance and Anger rating for sounds from BEST corpus

1.459). Low-rated sounds included Sound_2 ($M = 2.687, SD = 1.138$) and Sound_16 ($M = 2.625, SD = 1.310$).

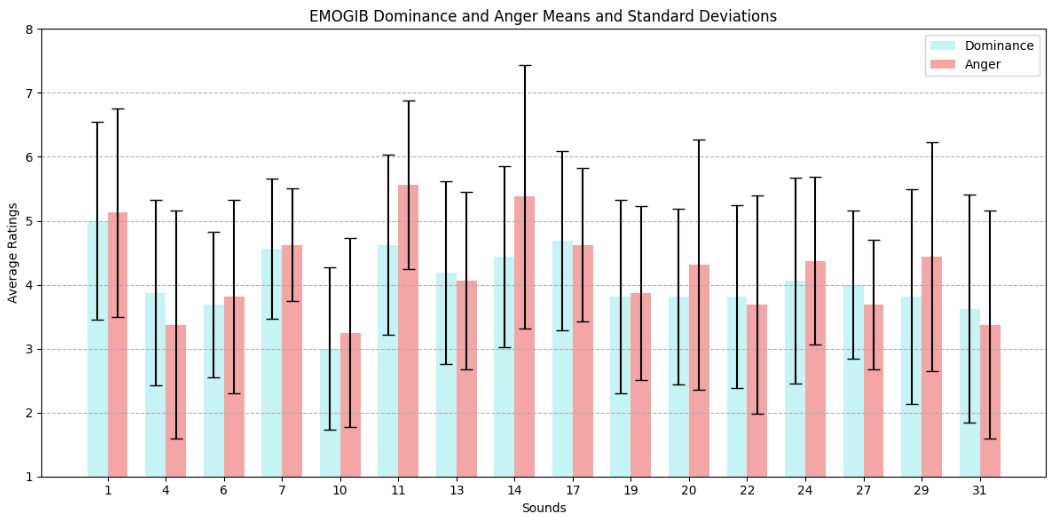


Figure 4.7: Dominance and Anger rating for sounds from EMOGIB corpus

Similarly, from the EMOGIB corpus (Figure 4.7), the top two sounds rated highest for Dominance were Sound_1 ($M = 5.000, SD = 1.549$) and Sound_17 ($M = 4.687, SD = 1.400$). For the medium-rated category, sounds close to a median of 3.938 were selected: Sound_27 ($M = 4.000, SD = 1.154$) and Sound_4 ($M = 3.875, SD = 1.454$). Low-rated sounds included Sound_31 ($M = 3.625, SD = 1.784$) and Sound_10 ($M = 3.000, SD = 1.264$).

Overall, these findings indicate that sounds (SFUs) from the EMOGIB corpus were perceived as more Dominant compared to those from the BEST corpus based on these participant ratings.

4.4 Discussion

This pre-study of the thesis was aimed at identifying SFUs to be selected for further examination in the main study in Chapter 5. The selection process involved the assessment of SFUs by 16 participants in a small within-subject design study, majorly to identify their perception of Dominance. This was done by calculating the mean ratings for Dominance expression within the EMOGIB and BEST corpora, the top two sounds in the high, medium, and low dominance categories are determined. Consequently, a total of six SFUs from each corpus are chosen for further analysis in the subsequent main study.

Moreover, the pre-study also explores the correlation between SFUs for their perception of Anger and Dominance. The results indicated a positive correlation between the perceived levels of dominance and anger, as reported by the $n = 16$ participants.

The limited sample size of $n = 16$ participants raises concerns regarding the robustness of the conclusions drawn from the pre-study. Additionally, certain demographic questions, such as those about the duration of residence in the country of origin and the country of residence, are deemed irrelevant and are thus excluded from the main-study.

Furthermore, the absence of open-ended questions accompanying each sound for participants to provide additional feedback is identified as a limitation, which is addressed in the main study. Future research could explore the underlying factors, one's culture influencing the perception of SFUs by utilizing a larger sample size. Investigating the broader implications of these findings in cross-cultural emotions could be valuable in the future in H-HRI studies.

Chapter 5

Main-Study

This chapter describes part two of this thesis: the main study, which was conducted using an online survey. The chapter covers the main study's purpose, design, measurement scales, analysis, and results.

5.1 Goal

The primary objective of the main study was to evaluate the extent to which SFUs from the EMOGIB and BEST databases can signal Urgency for a hospital delivery robot amongst the two cultures (Indian and Dutch). This involved examining the potential impact of Databases and Culture individually as the main effects and their impact together (interaction effect) on the perception of Urgency. Additionally, the study explored the relationship between perceived Dominance and the associated perceptions of the Urgency of these SFUs.

5.2 Study design

The main study uses a within-subject design, which was also used by the pre-study in Chapter 4, wherein each participant is exposed to the same sounds. As mentioned in the main study, 12 of the selected sounds based on their Dominance ratings from the pre-study were chosen for ease of evaluation, and the sequence of these stimuli is randomized to mitigate potential order effects.

The study was conducted using the Qualtrics survey platform [1]. The duration of the main study session spanned around 30–40 minutes and received ethical clearance from the ethics committee (EC-CIS) affiliated with the Faculty of Electrical Engineering, Mathematics, and Computer Science at the University of Twente.

[1]https://www.qualtrics.com/

The study was divided into three parts. Initially, participants were introduced to the survey, explained the notion of SFUs, and given explicit instructions, as illustrated in Figure B.1 and Figure B.2 in Appendix B. All participants had to sign the online consent form, as seen in Figure B.3, and only after that could they participate in the study. An image of a Harmony robot in a hospital lobby was used along with a human in front of the robot to serve as a visual aid to immerse participants in the contextual framework by whom, for whom, where and within which the auditory stimuli will be utilized is discussed in Section 5.2.1 and Measurements are discussed in Section 5.2.2.

Subsequently, after finishing the rating of the audio stimuli, participants went to part two of the study which is mentioned in Appendix B Figure B.6 and B.7 which are outside the scope of this study.

Lastly, participants were engaged in a demographic questionnaire to understand their demographics as a sample. The demographic questionnaire encompasses variables such as age, gender, national culture identification based on their nationality, cultural association of participants and prior exposure to robots. Detailed representations of the demographic survey can be found in Appendix A in figure A.4, figure A.5 and figure A.6.

5.2.1 Stimuli

From the pre-study as referred to in Chapter 4, twelve sound stimuli, six each from the EMOGIB and BEST corpora, were selected. Participants could play the sounds from their respective devices. These audio files were evaluated using the Qualtrics survey platform, where the sounds were complemented with a visual illustration of the Harmony robot in a hospital lobby (refer to the figure 5.1. This visual was created using text-to-image free AI software and edited using Canva [2].

The selection was divided into three categories: high, medium, and low urgency, each featuring two sounds from EMOGIB and BEST. These categories were strategically chosen to minimize bias.

5.2.2 Measurements

The main study utilized four scales of subjective measurement to evaluate the perception of twelve SFUs selected from the Pre-study from the BEST and EMOGIB corpora (refer to chapter 4 and section 4.2).

[2]https://www.canva.com/en_gb/

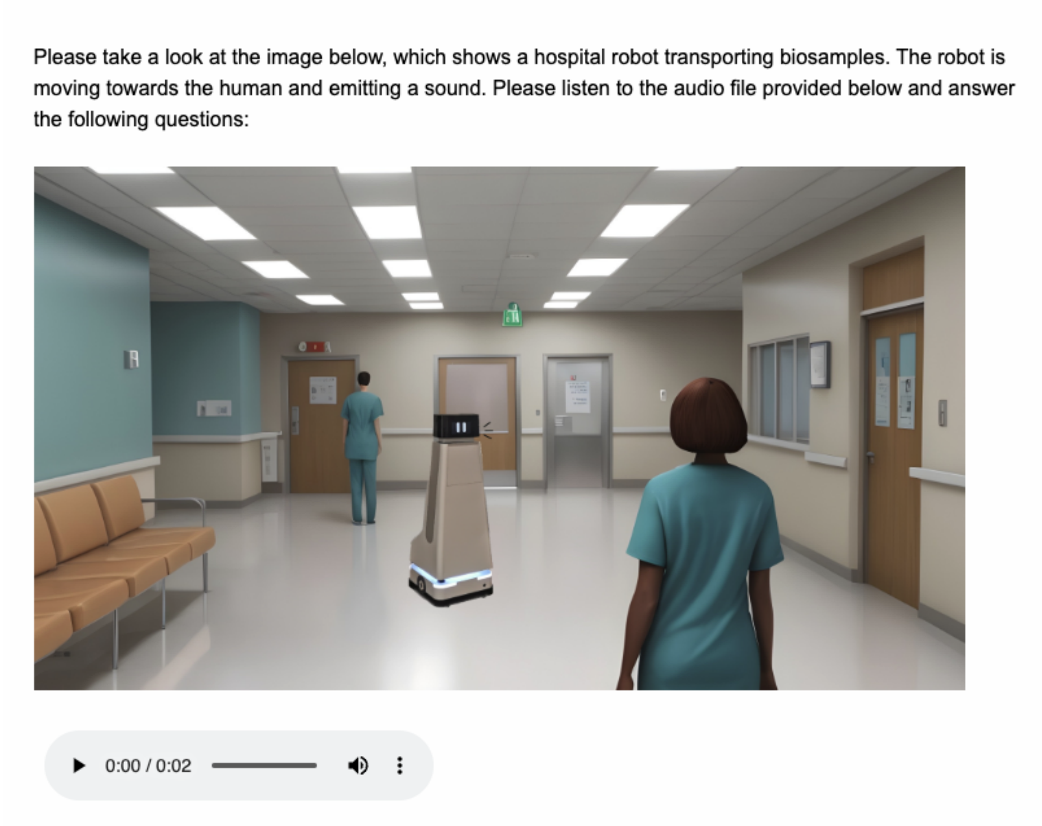
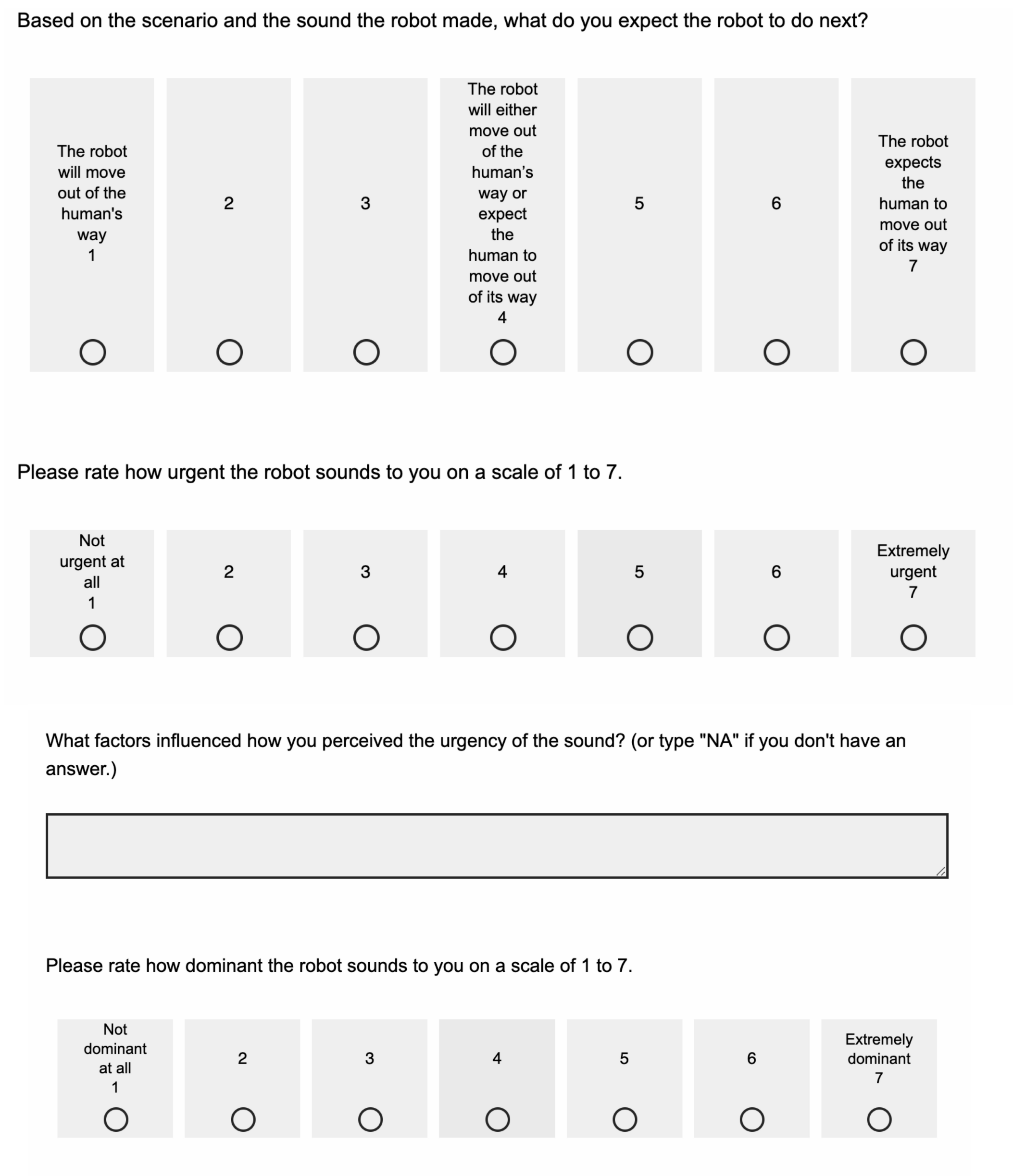


Figure 5.1: Stimuli showed to participants in the Main Study

Urgency Scales

In this study, two urgency scales were used: one is the main scale, also referred to as "Urgency_Scale1", and another one was from a study by Ottoman et al [2] referred to as 'Urgency_Scale2'. The objective was to assess the perceived urgency of the SFUs in 'Urgency_Scale1' and to ask the participants *"Please rate how urgent the robot sounds to you on a scale of 1 to 7"* using a Likert scale ranging from 1 to 7 (Refer to Figure 5.2). A score of 1 was the lowest score for Urgency, labelled "Not Urgent at all" and 7 was the highest, labelled "Extremely Urgent".

Urgency_Scale2 was adopted from Orthmann et al. [2]'s study, which was published during the later stages of this study before the experiment design phase. This adoption was done to support the findings from 'Urgency_Scale1' and to confirm that the intention was conveyed. It will not be utilized in the main analysis (Section 5.3.1) but only be used in descriptive analysis to find out how participants responded to the sound and to Urgency to support the 'Urgency_Scale1'. The scale asked the participants *"Based on the Scenario and the sound the robot made, what do you expect the robot to do next?"* and ranged from 1 to 7, with 1 labelled "The robot will move out of the human's way", 4 labelled "The robot will either move out of the human's way or expects the human to move out of its way", and 7 labelled as "The


Based on the scenario and the sound the robot made, what do you expect the robot to do next?

The robot will move out of the human's way 1	2	3	The robot will either move out of the human's way or expect the human to move out of its way 4	5	6	The robot expects the human to move out of its way 7

Please rate how urgent the robot sounds to you on a scale of 1 to 7.

Not urgent at all 1	2	3	4	5	6	Extremely urgent 7

What factors influenced how you perceived the urgency of the sound? (or type "NA" if you don't have an answer.)

Please rate how dominant the robot sounds to you on a scale of 1 to 7.

Not dominant at all 1	2	3	4	5	6	Extremely dominant 7

Figure 5.2: Qualtrics screenshot showing scales: Urgency_Scale1, Urgency_Scale2 [2] and Dominance

robot expects the human to move out of its way", which can be referred to in Figure 5.2.

Participants were also asked to write their responses on what factors influenced their perception of Urgency, as referred in Figure 5.2. This was done to ensure that the quantitative data was well supported with qualitative data and to evaluate it on communicating the intent of urgency using these SFUs.

Other Measurement Scales

The sounds were re-evaluated for Dominance in the main study *”Please rate how dominant the robot sounds to you on a scale of 1 to 7”* on the same Likert scale of 1 to 7, where 1 represented ”Not Dominant at all” and 7 represented ”Extremely Dominant”, for identifying the potential relationship between Dominance and Urgency of this SFUs.

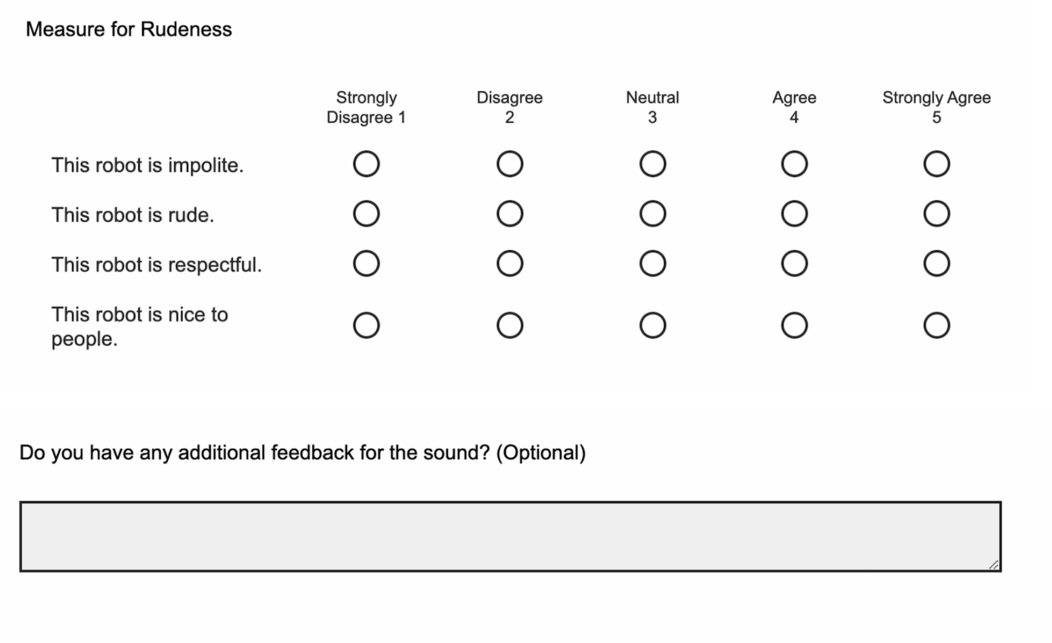


Figure 5.3: Qualtrics screenshot showing scales: Rudeness

Finally, the Rudeness (RUD-R) scale from the perceived social intelligence (PSI) scales [120] was used, which is particularly important for robots that will be present in social settings performing tasks around people. The scale assesses the robot’s ability to be a desirable social partner by displaying friendly, helpful, caring, and trustworthy behaviours, aiming to improve human-robot interactions, cooperation, and compliance. It was utilized to investigate the possibility that dominant SFUs, that might be perceived as Urgent could also be perceived as rude (refer to Section 2.4.3, ensuring their social acceptability and likability. The scale is a Likert scale ranging from 1 to 5, with 1 indicating ”Strongly Disagree”, 2 for ”Disagree”, 3 for ”Neither Agree nor Disagree”, 4 for ”Agree”, and 5 for ”Strongly Agree”. It applies to four conditions, which can be referred to in Figure 5.3.

Additionally, other measurements utilized in part two of the study can be found in the Appendix B Figure B.6 and B.7 are outside the scope of this study and will not be discussed in this thesis, but they may be discussed in other publications.

Apart from measurements for sound, part three of the study incorporated demographic questionnaires to gain a better understanding of the sample group, which is discussed in Section 5.2.3

5.2.3 Participants

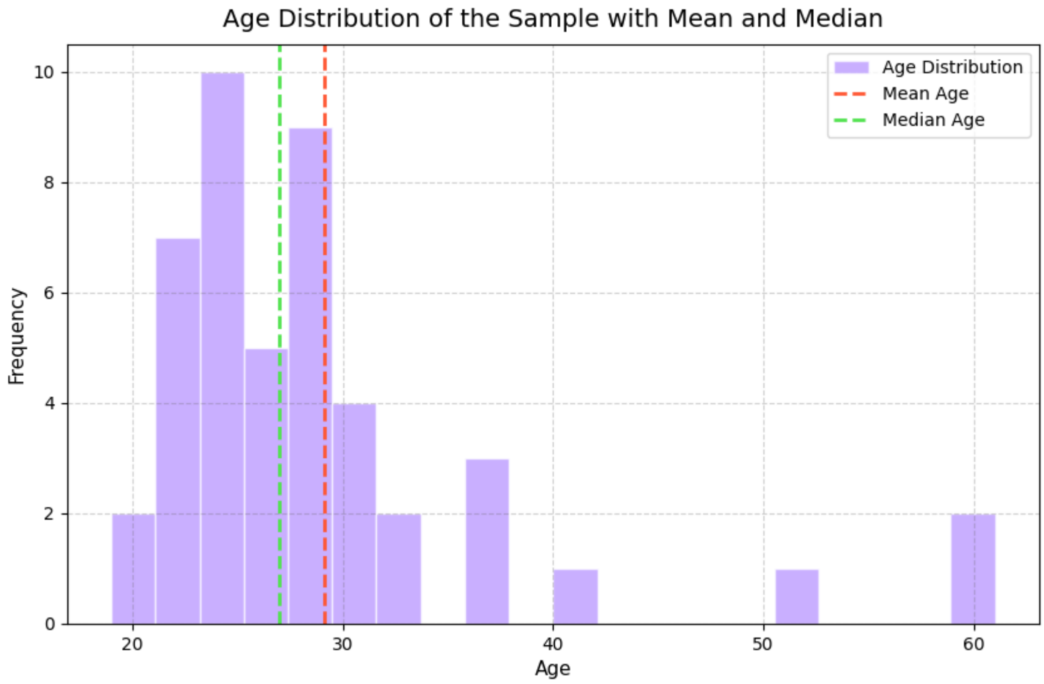


Figure 5.4: Age distribution of the Sample with Mean and Median

The sample resulted in a total of 49 participants, out of which three were excluded due to not giving consent, resulting in 46 participants, Dutch = 23, Indian = 23, separated based on their nationality for further analysis. The final sample ranged in age from 18 to 65 ($M = 29.13, SD = 8.83$), as detailed in Figure 5.4. Gender composition included Male = 25, Female = 19, non-binary = 1, prefer not-to-say = 1 and self-describe = 0, as shown in Figure 5.5.

Participants were asked about their nationality to identify their national culture (Figure B.4), which resulted in an equal representation of Indian and Dutch nationalities. The majority 76.1% resided in the Netherlands with only 19.6% in India and the rest 2.2% in the USA and Germany, as referred to in Figure 5.6. Since Dutch nationals can hold multiple nationalities, it was anticipated that some participants might have more than one cultural association, as illustrated in Figure B.5. Therefore, participants were asked about their cultural associations. The results showed that 50% solely associated themselves with Indian culture, with 41.3% associating themselves with Dutch culture and the rest 4.3% had their association to both the cultures Indian and Dutch or Dutch and other cultures same 4.3%, as referred to in Figure 5.6.

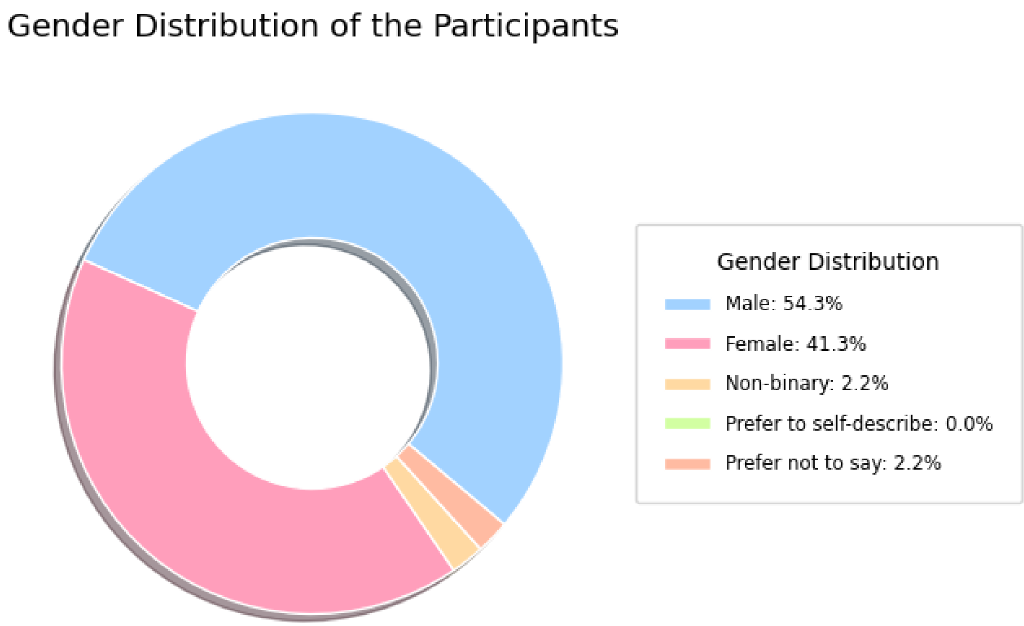
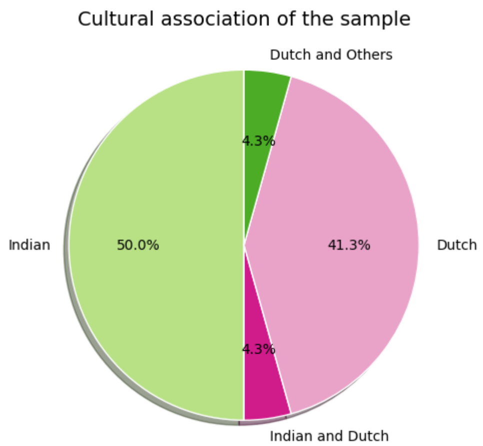


Figure 5.5: Gender distribution of the sample in the Main-study



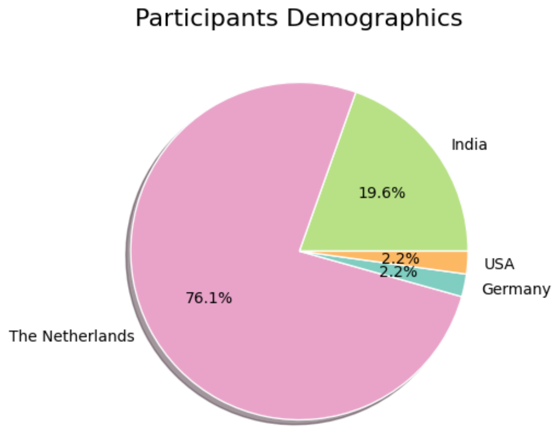


Figure 5.6: Cultural association as well as demographics of the Participants

Participants were also asked about their familiarity with robots (Figure B.5), where 6.5% were "Not familiar at all" with robots and 15.2% were "Very familiar" with robots, statistics of which can be referred to in Figure 5.7. Robot familiarity had different origins, the codes and code groups and their groundedness can be viewed from the table B.5 in Appendix B. Participants interacted with robots in various ways, such as watching movies, which was the most frequently given answer, owning robots at home, reading books, studying robotics in school, working on robot development projects, and engaging with robots in daily life and work environments.

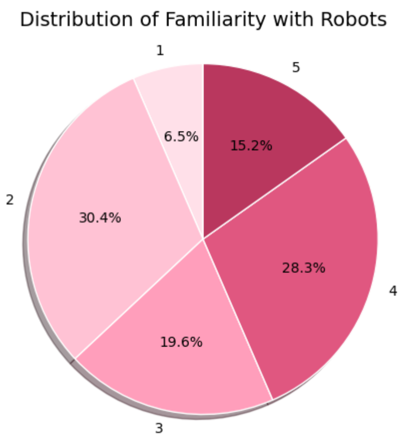


Figure 5.7: Participants' Level of Familiarity with Robots

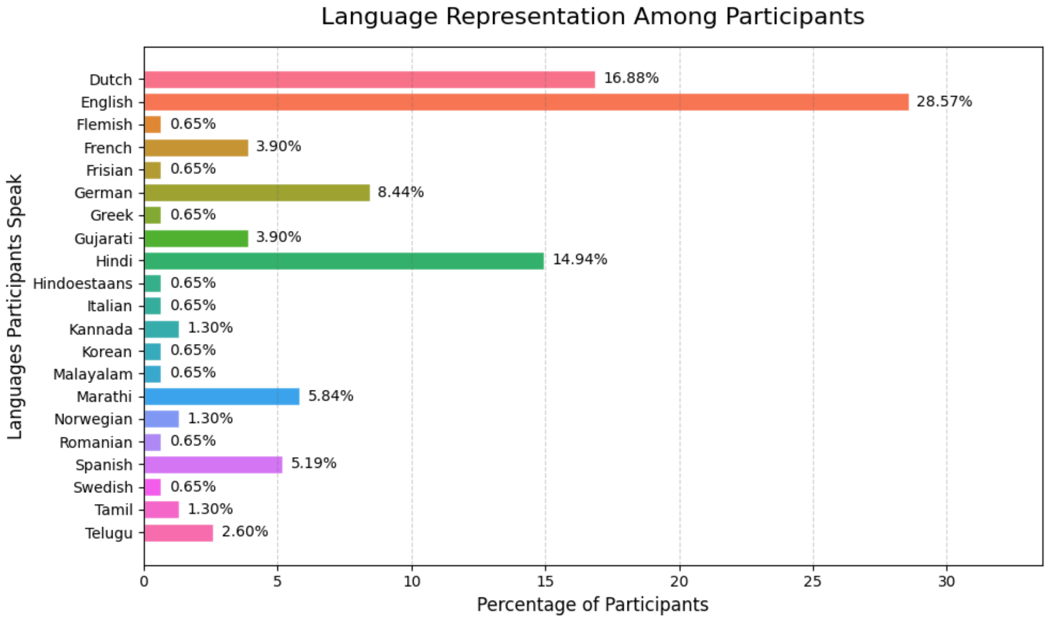


Figure 5.8: Language Representation Among Participants

When inquired regarding the languages they speak, the participants reported proficiency in a variety of primary languages *"English, Dutch, Hindi, Spanish, German"*. 34.78% of participants were bilingual and 65.22% were multilingual and fluent in additional languages such as *"English, Swedish, Norwegian, Hindi, Romanian, Malayalam, Tamil, Kannada, Telugu, Spanish, Marathi, Kannada, Gujarati, Hindoestaans, German, Korean, Frisian, Greek, Italian, and French"*. Figure 5.8 shows our participants' linguistic diversity. This multilingualism provides a good framework for researching the perceptions of urgency when using SFUs for delivery robots in the context of public spaces like hospitals, which host multicultural users.

5.3 Analysis and Results

In descriptive statistics for Dutch and Indian participants separate based on national culture across the BEST and EMOGIB databases, several key trends emerged regarding the perceptions of Urgency, Dominance, and Rudeness.

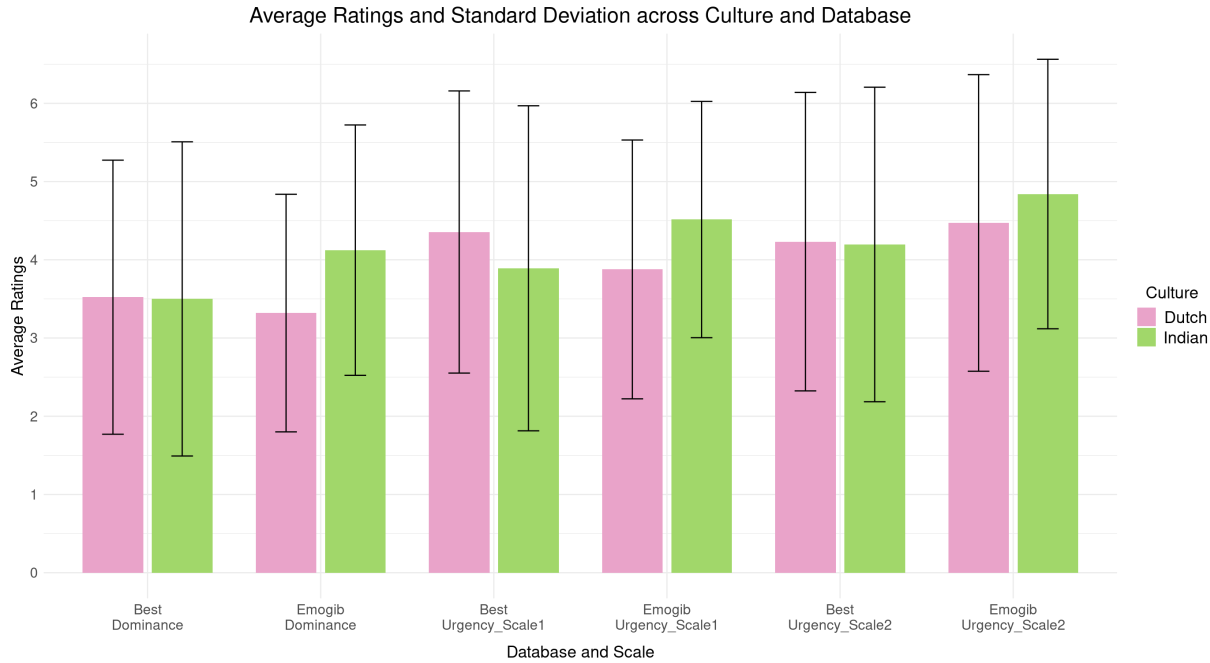


Figure 5.9: Average Ratings of Urgency_Scale1, Urgency_Scale2 and Dominance scales across Culture and Database

For Urgency_Scale1 as seen from Figure 5.9 and Figure 5.10, Dutch participants rated Urgency higher in the BEST database (M $= 4.36$, SD $= 1.80$, Mdn $= 5$, IQR $=$ 3) than in EMOGIB database (M $= 3.88$, SD $= 1.65$, Mdn $= 4$, IQR $= 2$). In contrast, Indian participants showed an opposite trend, with higher Urgency reported in the EMOGIB database (M $= 4.51$, SD $= 1.51$, Mdn $= 5$, IQR $= 2.75$) compared to the BEST database (M $= 3.89$, SD $= 2.08$, Mdn $= 4$, IQR $= 4$). The larger IQR for the Indian BEST dataset suggests greater variability in their responses. This suggests that although Dutch participants demonstrate higher Urgency in the BEST database, Indian participants perceive greater Urgency in the EMOGIB database, indicating a difference in sensitivity to contextual factors across cultures.

For Urgency_Scale2, Dutch participants perceived similar Urgency ratings across both databases, with slightly higher ratings in the EMOGIB database ($M = 4.47$, $SD = 1.90$, $Mdn = 5$, $IQR = 3$) compared to the BEST database ($M = 4.23$, $SD = 1.91$, $Mdn = 4$, $IQR = 3$). Indian participants also perceive slightly greater urgency

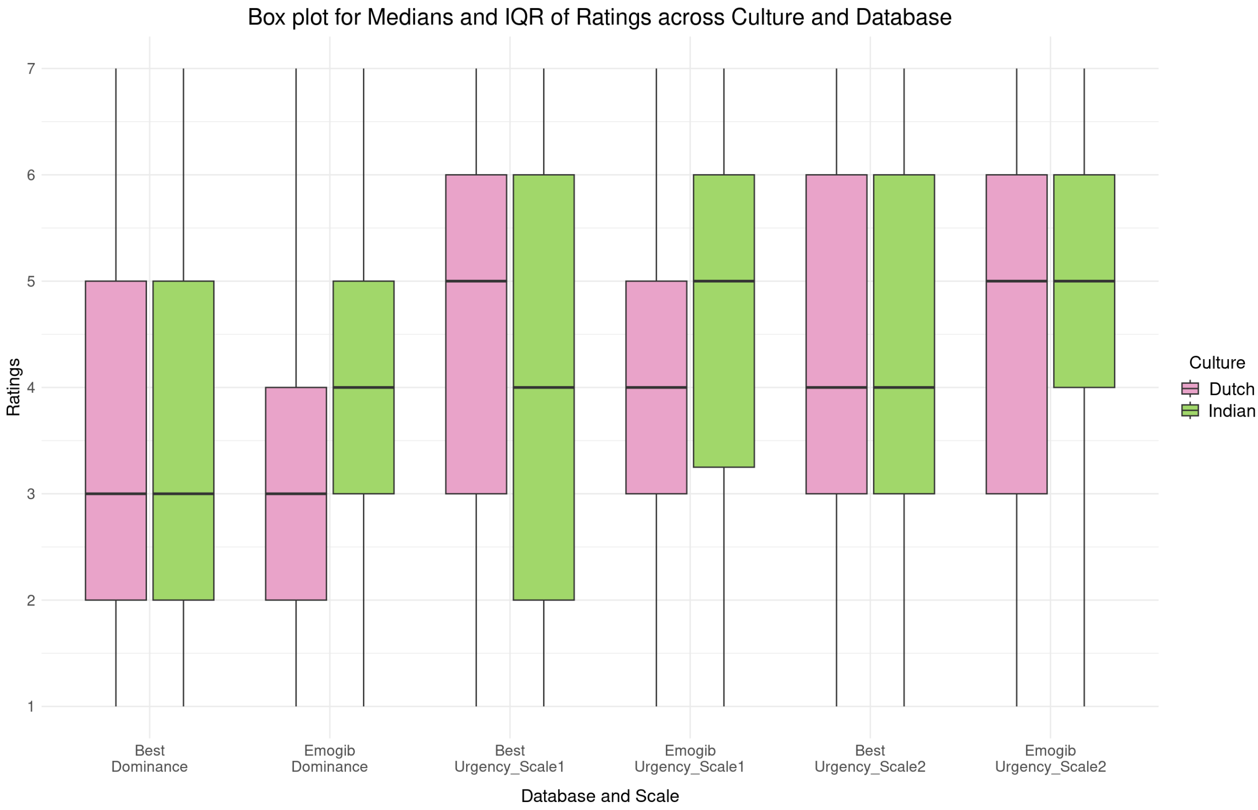


Figure 5.10: Box plot for Medians and IQR of Ratings across Culture and Database

in the EMOGIB database ($M = 4.84$, $SD = 1.72$, $Mdn = 5$, $IQR = 2$) relative to the BEST database ($M = 4.20$, $SD = 2.01$, $Mdn = 4$, $IQR = 3$). This indicates that although database context influences Urgency ratings for both cultures, Indian participants are slightly more sensitive to these contextual differences compared to Dutch participants.

In terms of Dominance Ratings, Dutch participants reported only a slightly higher Dominance in the BEST database ($M = 3.52$, $SD = 1.75$, $Mdn = 3$, $IQR = 3$) compared to the EMOGIB database ($M = 3.32$, $SD = 1.52$, $Mdn = 3$, $IQR = 2$). Although there is a difference, the magnitude of this difference is relatively small, and both medians are identical. In contrast, Indian participants rated Dominance significantly higher in the EMOGIB database ($M = 4.12$, $SD = 1.60$, $Mdn = 4$, $IQR = 2$) compared to the BEST database ($M = 3.50$, $SD = 2.01$, $Mdn = 3$, $IQR = 3$). The higher Dominance ratings in the EMOGIB database by Indian participants, combined with the smaller IQR suggest a stronger and more consistent perception of Dominance in this context. These results indicate that database context has a more pronounced effect on Dominance perceptions among Indian participants compared to Dutch participants than on Urgency_Scale1 and Urgency_Scale2.

For Rudeness Ratings from Figure 5.11 and Figure 5.12, both Dutch and Indian participants reported relatively consistent rudeness ratings across the databases.

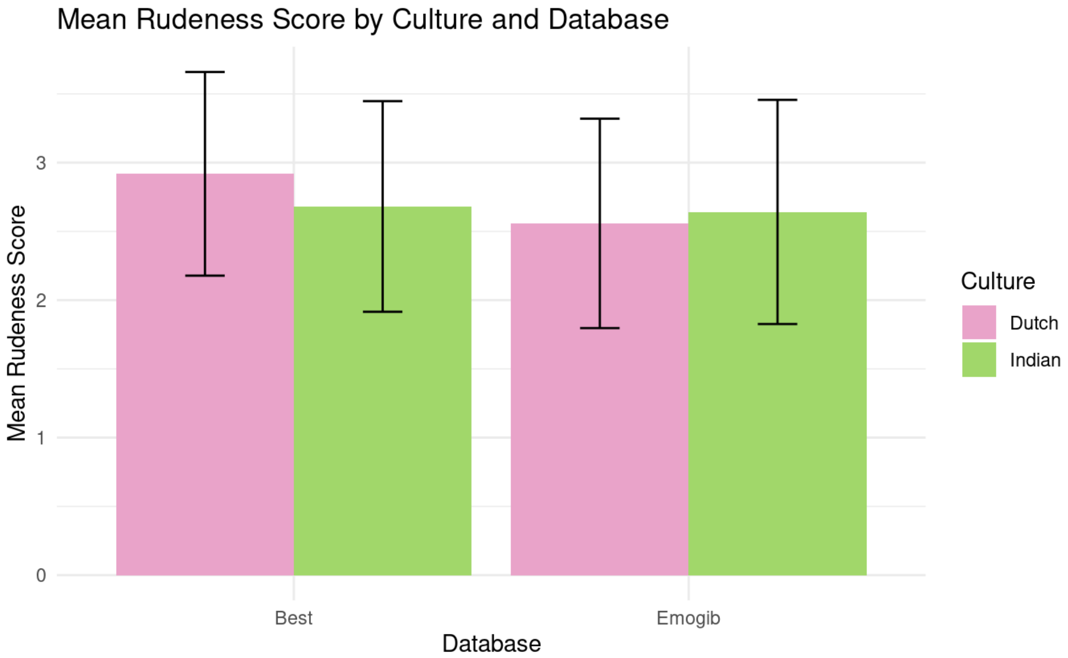


Figure 5.11: Average Ratings of Rudeness scales across Culture and Database

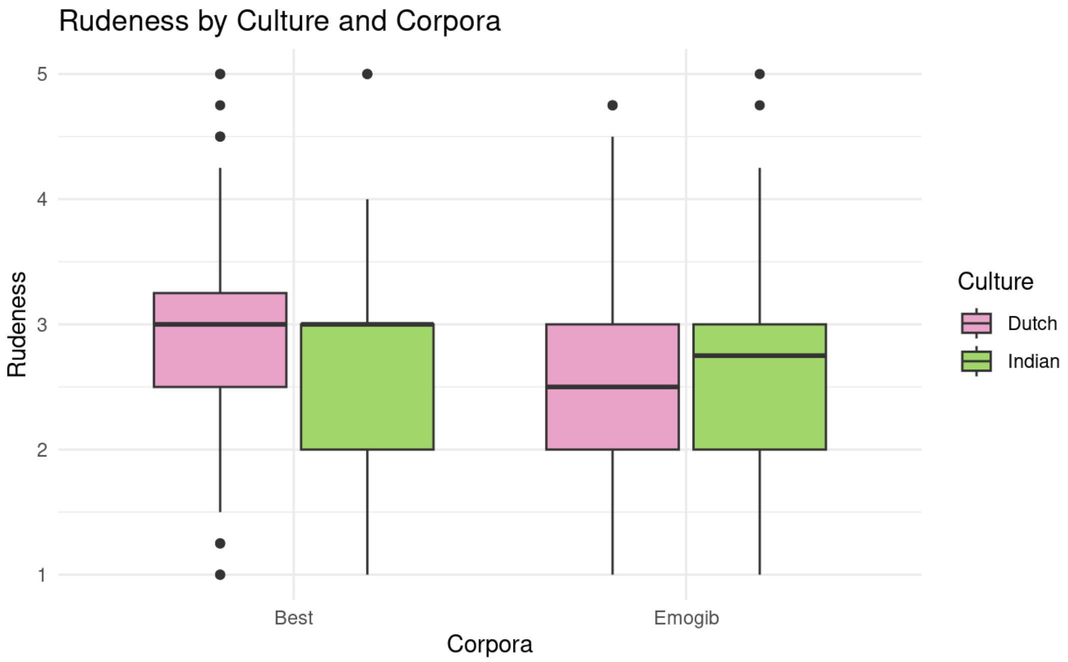


Figure 5.12: Boxplot of Rudeness Ratings across culture and database

Dutch participants rated rudeness slightly higher in the BEST database ($M = 2.92$, $SD = 0.74$, $Mdn = 3$, $IQR = 0.75$) compared to the EMOGIB database ($M = 2.56$, $SD = 0.76$, $Mdn = 2.5$, $IQR = 1$). Indian participants exhibited minimal variation between databases ($M = 2.68$, $SD = 0.77$, $Mdn = 3$, $IQR = 1$ in BEST and $M = 2.64$, $SD = 0.82$, $Mdn = 2.75$, $IQR = 1$ in EMOGIB). This suggests that perceptions of Rudeness are relatively stable and overall low across both cultural backgrounds and databases.

In conclusion, these descriptive statistics underscore the importance of considering cultural and database factors when interpreting emotional and behavioural evaluations. Although Urgency and Dominance perceptions were influenced by database context, with Indian participants showing more pronounced effects, perceptions of Rudeness remain relatively stable across different contexts and cultures. The scale were explored further in Section 5.3.2 and the impact of cultural background and Database in Section 5.3.1.

5.3.1 Effects of Culture and Database on Urgency

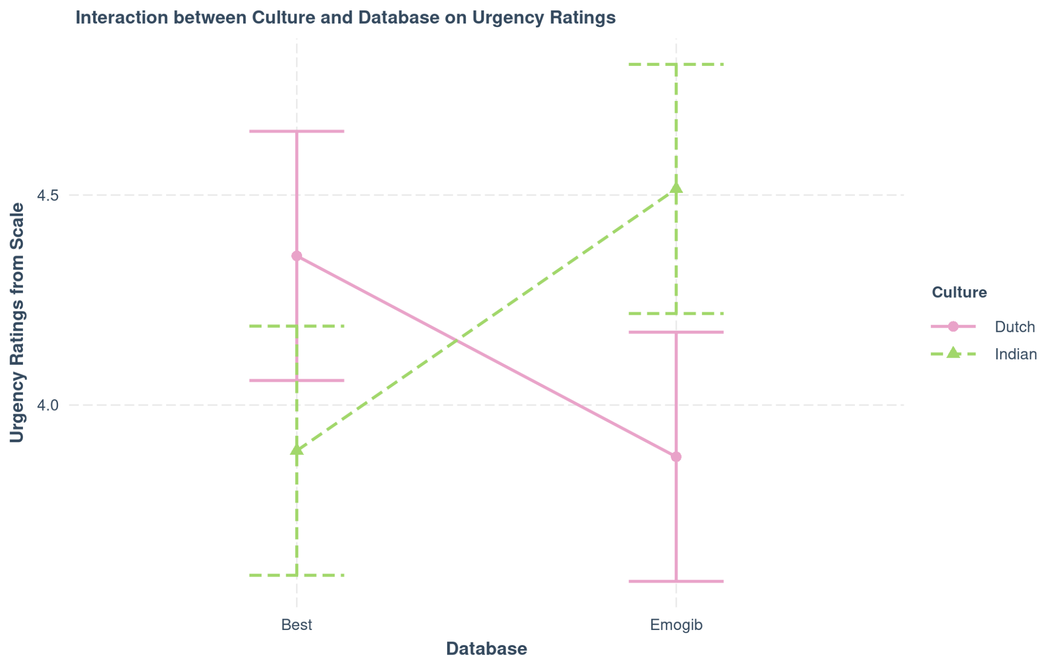


Figure 5.13: Interaction plot between the Fixed Factors: Database and Culture for Urgency_Scale1

The interaction plot as referred to in Figure 5.13 shows a cross-over interaction effect between fixed factors, i.e., Culture and Database for Urgency_Scale1. The plot suggests that for Dutch participants, urgency ratings decreased from the ”BEST” to the ”EMOGIB” database, whereas for Indian participants, the opposite trend was observed. This leads to further analysis of these factors.

A mixed-effects model, Aligned Rank Transform (ART) ANOVA [121] from ARTool package in R was utilized to test this interaction effect’s significance as well as the main factors effect. This method is particularly useful for handling non-parametric data that doesn't meet normality assumptions for factorial designs. The model used in this study was:

art(Urgency_Scale1 $\sim$ Culture$\times$Database$+(1|$Participant_ID$)+(1|$Sound_ID$)$, data = df1)

The model includes random effects for participants and sound, which can account for potential variability within these groups. The alignment process in the Aligned Rank Transform (ART) ANOVA was verified by ensuring that the column sums of the aligned responses were approximately zero for each factor (Culture = 0, Database = 0, Culture: Database = 0). This verification confirms that the data transformation was performed correctly, allowing for the subsequent application of traditional ANOVA to the transformed data. The results of the ART model ANOVA are summarized in Table 5.1.

Table 5.1: Analysis of Variance of Aligned Rank Transformed Data (ART ANOVA)

Effect	F	df	df_{res}	p	Partial η^2
Culture	0.027	1	44	.870	.0006
Database	0.015	1	10	.904	.0015
Culture:Database	16.488	1	494	<.001	.0323

The ART ANOVA results revealed that neither the main effects of Culture ($F(1,44) = 0.027$, $p = .870$, partial $\eta^2 = .0006$) nor the Database ($F(1,10) = 0.015$, $p = .904$) partial $\eta^2 = .0015$ were significant. This indicates that these factors alone do not significantly influence urgency ratings with effect sizes for Culture and Database being very small. However, the interaction between Culture and Database was significant ($F(1,494) = 16.488$, $p < .001$, partial $\eta^2 = .0323$), suggesting that the perception of urgency is influenced by the combination of culture and database.

The observed interaction effect ($F(1,494) = 16.488$, $p < .001$, partial $\eta^2 = .0323$) suggests a small to medium effect size. This cross-over interaction effect was also illustrated in the plot 5.13, demonstrating that the relationship between the one-factor Culture and the dependent variable of Urgency changes as a function of the level of the second factor, Database. In other words, the influence of Culture on the perception of urgency differs for different Databases contexts.

To determine statistically significant interaction effects within the levels of the factors, i.e., Indian and Dutch for Culture and EMOGIB and BEST for Database, a post-hoc test to make pairwise comparisons using the `art.con()` function from the same ARTool package was performed. The results of the contrasts are presented in Table 5.2.

The ART ANOVA revealed a significant interaction effect, however, it is to be noted that none of the pairwise comparisons between the various combinations of

Table 5.2: Pairwise comparisons for the Interaction between Culture and Database

Contrast	Estimate	SE	*df*	t-ratio	p
Dutch, BEST - Dutch, EMOGIB	41.80	40.3	11.7	1.038	.731
Dutch, BEST - Indian, BEST	38.95	24.2	70.4	1.610	.380
Dutch, BEST - Indian, EMOGIB	-12.56	44.3	16.8	-0.284	.992
Dutch, EMOGIB - Indian, BEST	-2.85	44.3	16.8	-0.064	.999
Dutch, EMOGIB - Indian, EMOGIB	-54.36	24.2	70.4	-2.247	.121
Indian, BEST - Indian, EMOGIB	-51.51	40.3	11.7	-1.279	.593

Culture and Database (a family of 4 estimates) reached statistical significance ($p > .05$) after applying Tukey's adjustment for multiple comparisons.

This outcome suggests that, although the interaction was significant at the ART ANOVA level($F(1, 494) = 16.488$, $p < .001$, $\eta^2 = .0323$), the specific pairwise comparisons failed to show any statistically significant differences amongst the groups. This could be due to the complexity of the interaction or to a possibly insufficient sample size, which may limit the identification of these differences at a finer level.

The compact letter display (Table 5.3) further confirmed that all groups shared the same letter. This further confirmed that no significant differences exist between any pairs of groups. Therefore, it is important to consider both cultural factors and database information when analyzing urgency ratings, rather than examining them separately.

Table 5.3: Compact letter display of Groups

Group	**Letter**	**MonoLetter**
Dutch,BEST	a	a
Dutch,EMOGIB	a	a
Indian,BEST	a	a
Indian,EMOGIB	a	a

The explanatory power of the model, or goodness of fit (which includes the fixed effects of Culture, Database, and their interaction, as well as random effects for Participant_ID and Sound_ID) was further assessed using Efron's pseudo-R-squared and the R-squared from linear model that does not include random effects. Both values were approximately 0.025, which indicates that the models explain only about 2.47% of the variance in Urgency (Urgency_Scale1). These low R-squared values indicate that, although the interaction between Culture and Database was statistically significant, it only explains a small portion of the variability in Urgency ratings. This highlights the complexity of factors influencing Urgency perception and implies that other variables, not included in the model, might also play a significant role.

The thematic analysis of the qualitative data was also conducted for the question of *What factors influenced how you perceived the Urgency of the sound?*, which can be referred from Figure 5.2. It reveals insights into how Indian and Dutch participants perceived the acoustic properties that influenced their perception of urgency. The results are summarized in Table 5.4

Table 5.4: Thematic Analysis of Factors Influencing Urgency Perception

Group	Theme	Occurrence	Example Quotations
Indian - EMOGIB	Pitch	12	”High pitch but low pace of speech”, ”Too short to be urgent”
	Tone	10	”Tone of the voice”, ”Sounded rude/shouting”
	Urgency Perception	8	”There was not much urgency in the tone”, ”Sounds like a request to move”
Dutch - EMOGIB	Pitch	6	”High pitch, feminine sounding”, ”The quick tonal changes make it sound distressed”
	Urgency Perception	10	”It sounds a bit urgent, but not extremely”, ”The robot sounds distressed”
	Tone	8	”The tone of voice sounds polite and empathetic”, ”Sounds like a nagging person”
Indian - BEST	Alarm-like sounds	8	”Sounds like an emergency siren”, ”This sounds like an alarm”
	Pitch	6	”High pitch and pace”, ”Rising pitch”
	Urgency Perception	8	”Loud increase in frequency tells you to move away”, ”Sounds highly urgent”
Dutch - BEST	Alarm-like sounds	7	”Sounds like an air alarm”, ”Alarm sound, watch out!”
	Pitch	5	”Increasing pitch”, ”Very loud sound”
	Urgency Perception	6	”Persistent, increasing in anger”, ”I would immediately jump out of the way for this robot”

Both Indian and Dutch participants recognized the intent of communication and the acoustic properties of this SFUs for conveying urgency. This is indicated by participants from both Indian and Dutch cultures, who find them as high-pitched, similar to an alarm or an emergency siren and associated with urgency. SFUs from the BEST database, participants from both cultural backgrounds found that sounds resembling emergency sirens or alarms were particularly attention-grabbing and perceived as highly urgent. SFUs from EMOGIB were perceived as high-pitched by both participants. However, Indian participants listening for the EMOGIB sounds indicated that the tone was frequently described as rude or shouting, contributing to a lower perception of urgency. In contrast, Dutch participants found the tone more polite and empathetic, though some also described it as nagging, which affected their urgency perception differently.

For the EMOGIB corpus, both Indian and Dutch participants identified Pitch, Tone, and Urgency Perception as crucial factors. Indian participants described the sound as having a ”High pitch but low pace of speech,” noting it was ”Too short to be urgent,” with the tone perceived as ”rude/shouting.” They often felt ”not much urgency in the tone” and that it ”sounded like a request to move.” Dutch participants similarly noted the high pitch and quick tonal changes, which made the sound ”distressed,” and characterized the tone as ”polite and empathetic” or ”nagging.” They remarked that it ”sounds a bit urgent, but not extremely.”

In the BEST group, both Indian and Dutch participants highlighted Alarm-like sounds, Pitch, and Urgency Perception. Indian participants frequently described alarm-like sounds as ”emergency sirens” or ”an alarm,” with a ”high pitch and pace” and ”rising pitch,” which conveyed ”high urgency.” Dutch participants described alarm-like sounds similarly, noting them as ”air alarms” or ”alarm sounds,” with ”increasing pitch” and ”very loud sound.” They perceived the sound as ”persistent, increasing in anger,” indicating they ”would immediately jump out of the way for this robot.”

Overall, the thematic analysis indicates that both Indian and Dutch participants perceived urgency through similar acoustic properties such as pitch and tone, which influenced their perception of urgency communication as it sounded similar to alarm-like sounds, although their specific descriptions varied.

5.3.2 Relation between Scales

The relation between the scales when using these SFUs was investigated using Kendall’s tau-b correlation. Correlation tests were for combinations of scales which can be referred to from Table 5.5.

Table 5.5: Kendall’s Tau Correlation Between Rudeness, Dominance, and Urgency Ratings

Correlation Combination	Kendall’s τ	z-value	p-value
Dominance vs Urgency_Scale1	0.615	18.738	$< 2.2 \times 10^{-16}$
Rudeness vs Urgency_Scale1	0.174	5.411	6.268×10^{-8}
Dominance vs Rudeness	0.302	9.384	$< 2.2 \times 10^{-16}$

Relation between Urgency_Scale and Dominance

The relation between Urgency_Scale1 and Dominance was evaluated using Kendall's τ correlation from Table 5.5. A strong positive correlation $\tau = 0.615, p < 0.001$ was found which is also indicated by the pink trend line in the plot in Figure 5.14, suggesting that SFUs which are perceived as high on Dominance are also perceived as high on Urgency, and vice-versa.

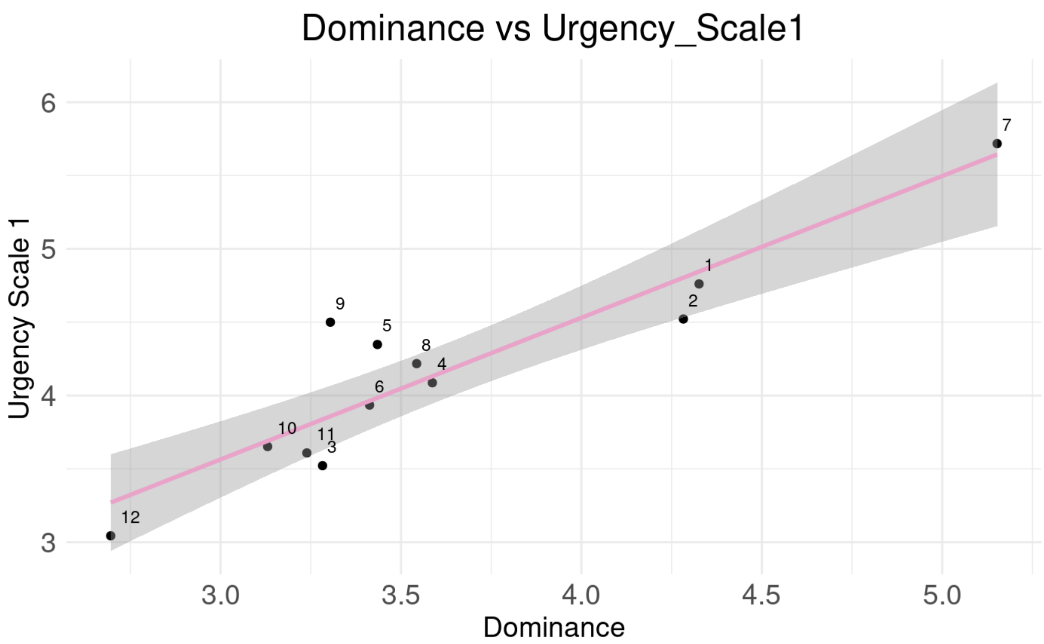


Figure 5.14: Scatter plot for Correlation between Dominance and Urgency_Scale1

Relation between Dominance and Rudeness

The findings from Table 5.5 and Scatter plot in Figure 5.15 suggest a moderately positive correlation between Dominance and Rudeness ($\tau = 0.302$) indicating that as perceived Dominance of SFUs increases, so does its perceived Rudeness. This correlation is statistically significant but not strong, which indicates that Dominance may not always predict Rudeness. The graph suggests that specific data points, sounds for example Sound_4 is high on Rudeness but lower on Dominance, and Sound_1 is higher on Dominance but lower on Rudeness, however, Sound_7 is high on both scales. This highlights the variability within the data. The clusters of points around the trend line suggest a consistent relationship between the two scales, although the correlation is moderate.

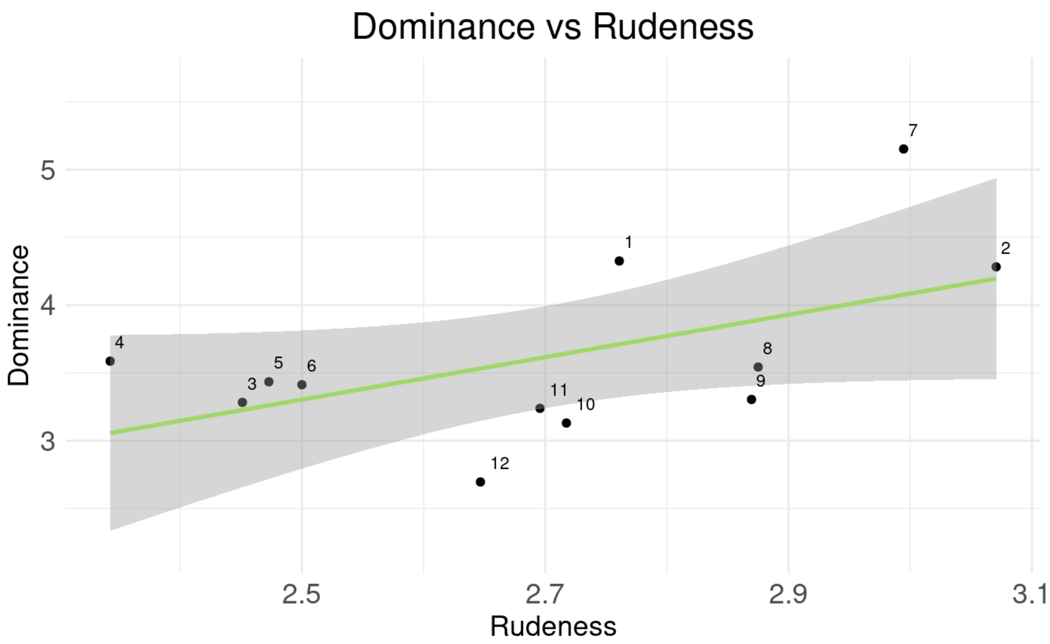


Figure 5.15: Scatter plot for Correlation between Dominance and Rudeness

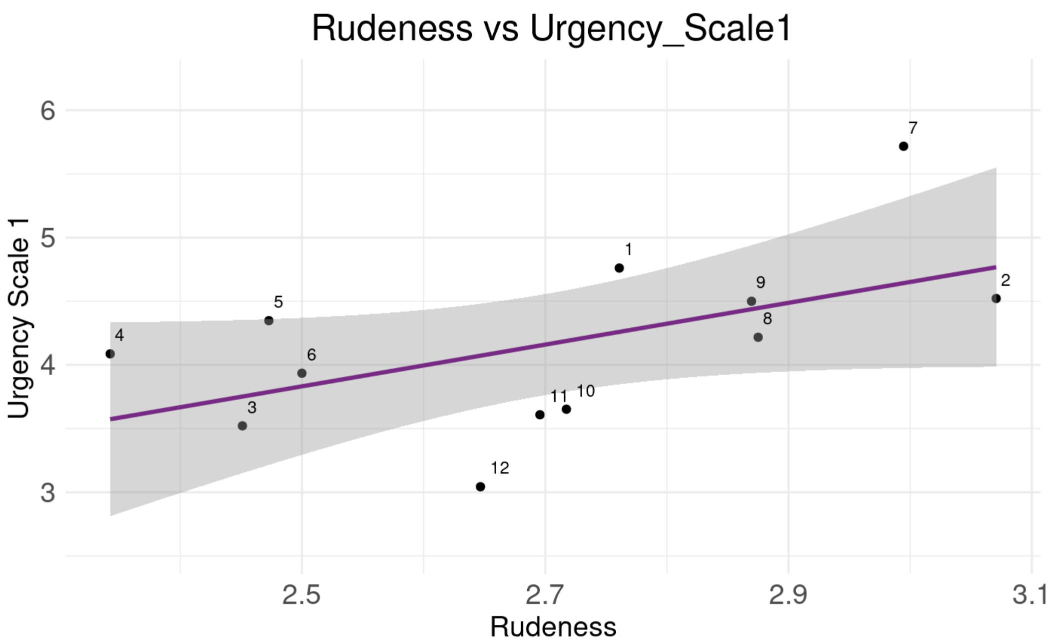


Figure 5.16: Scatter plot for Correlation between Rudeness and Urgency_Scale1

Relation between Urgency_Scale1 and Rudeness

The weaker correlation between Urgency_Scale1 and Rudeness ($\tau = 0.174$) from Table 5.5 indicates only a slight association between these two measurement scales. This suggests that Urgency_Scale1 does not strongly predict Rudeness, and vice versa. For example, in Figure 5.16 Sound_7 is high on Urgency_Scale1 is $M =$

5.717 and moderate on Rudeness is $M = 2.994$. However, Sound_4 is moderate to high on Urgency_Scale1 ($M = 4.087$) but low on Rudeness ($M = 2.342$). Similarly, Sound_5 is high on Urgency_Scale1 $M = 4.348$ but moderate to low on Rudeness $M = 2.473$ however Sound_2 is high on Urgency_Scale1 is also high on Rudeness $M = 3.070$. This correlation between Rudeness and Urgency_Scale1 is significantly less pronounced than between Dominance and Rudeness.

These findings suggest there can be urgent sounds without being rude; however, these sounds might still exhibit some level of Dominance.

5.3.3 Secondary Analysis

In the previous sections, participants were classified solely by their nationality as either Dutch or Indian, allowing for a direct comparison between the two groups. However, as mentioned in Section 5.2.3 and illustrated in Figure 5.6, indicated that a small percentage of individuals identified with multiple cultural norms.

A secondary analysis was conducted to explore whether accounting for these multiple cultural associations might yield different results. The Aligned Rank Transform (ART) ANOVA was performed, incorporating broader cultural associations as a fixed factor. The results of this analysis are displayed in Table 5.6.

Table 5.6: Analysis of Variance of Aligned Rank Transformed Data for Secondary Analysis (ART ANOVA)

Effect	F	df	df_{res}	p	Partial η^2
Cultural_Association	2.205	3	518.55	.0865	0.013
Database	0.072	1	14.57	.792	0.005
Cultural_Association:Database	1.550	3	518.30	.201	0.009

The ART-ANOVA results indicated that neither the main effect of Cultural Association ($F(3, 518.55) = 2.205$, $p > .05$, partial $\eta^2 = 0.013$) nor the main effect of Database ($F(3, 518.55) = 2.205$, $p = .087$, partial $\eta^2 = 0.013$) was significant. The partial eta-squared value of partial $\eta^2 = 0.013$ indicates that cultural association accounts for only about 1.3% of the variance in Urgency ratings, which reflects a small effect size and Database accounts for 0.5% of the variance in Urgency ratings. Thus, the Urgency ratings were consistent across both databases, regardless of the participant's cultural associations.

The interaction between Cultural Association and Database was also not significant ($F(3, 518.30) = 1.550$, $p = .201$, partial $\eta^2 = 0.009$). Although cultural associations may influence perceptions of Urgency, their effect does not appear to significantly interact with the type of database.

In conclusion, given the lack of significant results, the Urgency ratings were consistent across both databases, regardless of the participant's cultural association. This suggests that neither cultural association nor the type of database used had a meaningful impact on how participants perceived the Urgency of the stimuli in this study. As the secondary analysis yielded no new insights, it will not be discussed further. Overall, these results suggest that SFUs can be reliably used in multicultural settings without concern for cultural bias affecting urgency ratings.

Chapter 6

Discussion

This study focused on exploring SFUs from BEST and EMOGIB corpora that are available as open-source resources for research. Drawing inspiration from everyday HHI, we sought to identify SFUs that could effectively convey urgency for the service robot [1] in a cross-culture setting. As discussed in Chapter 2 Section 2.4.3, Anger can be a prosocial behaviour to help attain goals and to convey Dominance (Section 2.4.3, and Dominant behaviour has the potential to communicate Urgency (Section 2.4.2).

The ultimate goal of this study was to determine if SFUs, could be universally perceived as urgent across different cultures, making them suitable for use in public settings like hospitals. We proposed several hypotheses to explore these dynamics and gathered data to test these assumptions. For understanding the effect we proposed: *H1: In a hospital delivery robot, the urgency perception of semantic-free utterances from BEST and EMOGIB will differ, regardless of participants' cultural backgrounds.*, *H2: In a hospital delivery robot, the urgency perception of semantic-free utterances will differ amongst the participants of India and the Netherlands, regardless of Corpora*, *H3: The interaction between the corpora (BEST vs. EMOGIB) and the cultural backgrounds (Indian vs. Dutch) will significantly affect the perception of urgency in a hospital delivery robot.*

To understand the relation between Dominance and Urgency we proposed: *H4: In a hospital delivery robot, the perception of dominance is positively correlated with urgency conveyed by SFUs across different cultures.*

6.1 Effect of Culture and Database on Urgency

The effect of culture, database, and their interaction on the population was evaluated using ART ANOVA and post hoc tests to investigate hypotheses H1, H2, and H3.

6.1.1 Effect of Database alone

The study examined the main effect of the EMOGIB and BEST databases on urgency perception, as proposed in H1. Findings from the quantitative analysis using the mixed-effects model (ART ANOVA) showed no significant impact on urgency due to the database factor alone, leading to the rejection of hypothesis H1. This indicates that the type of database alone does not significantly influence urgency perception.

Qualitative feedback from participants is mentioned in Table 5.4 in Appendix B indicates that both Dutch and Indian participants frequently mentioned perceived urgency in their responses. Participants mentioned that the tone of voice, pitch, and loudness were factors associated with higher urgency perceptions in both cultures. However, some participants from both cultures found the sounds to be less urgent, which was not linked to a specific database.

These variations in acoustic properties like pitch, loudness, and length of speech components, as discussed by Zhang et al. [65], reflect the speaker's emotional state, making it easier to identify basic emotions across cultures using vocal cues

These findings, discussed in Chapter 2 and Chapter 3, are supported by EMOGIB and BEST, which are comprised of GS and NLUs and have been evaluated for conveying affective communication and expressing basic emotions despite the absence of the comprehensible words [13], [16], [87]. The current study validates these in cross-cultural backgrounds.

Hence, it suggests that neither Indian nor Dutch participants consistently perceived a significant difference in urgency between BEST and EMOGIB rather, their perception was influenced by additional factors like pitch and tone.

6.1.2 Effect of Culture alone

Similarly, the main effect of Culture alone also showed no significant difference in Urgency perception between Indian and Dutch participants, rejecting Hypothesis H2.

The qualitative data supports this conclusion, as the basic acoustic properties indicated by the themes (refer Table 5.4 in Appendix B), such as pitch, tone, and loudness, were similarly perceived by both Indian and Dutch participants. Some participants also found these sounds "attention-grabbing," like alarms or sirens. This supports the notion that urgency is recognized across both individualist and collectivist cultures based on inference from the sample size and culture studied in this research. This viewpoint aligns with Ekman and Izard's [63], [64] work that basic emotions (anger, disgust, fear, happiness, sadness and surprise) in facial expressions are universally recognizable could also be recognizable in vocal expressions, at least for anger. The sense of urgency conveyed through these basic emotions is also universally (cross-culturally) recognizable in non-verbal vocal communication. This leads to the rejection of hypothesis H2, indicating that cultural background alone did not significantly influence the perception of urgency.

6.1.3 Effect of Culture and Database together

The findings for the interaction effect between Culture and Database suggested a significant effect, which was supported by partial η^2 indicating a meaningful interaction. This suggests that combining cultural background with the specific database significantly influences the perception of urgency.

Qualitative data showed mixed results (refer Table 5.4 in Appendix B, with Indian participants often referencing emotional cues like anger or politeness, while Dutch participants focused more on the clarity of the sound and its similarity to human-like requests or alarms. Dutch participants also found EMOGIB sounds less urgent than Indian participants, and the reverse trend is seen in BEST. Although these differences in cultural perceptions of urgency were observed, both groups displayed diverse responses, which does not strictly confirm the hypothesis of distinct cultural perceptions. This suggests that the interaction between cultural background and specific acoustic features can influence the perception of urgency and the notion that, although basic emotions are universally understood, their perceived intensity can vary depending on cultural background.

However, pairwise comparisons did not reveal significant differences between groups, indicating that although the interaction effect is significant, specific pairwise comparisons did not demonstrate substantial variations. The model fit (effect size) was low, and the pairwise comparisons suggested that the significance of this effect between specific groups was not substantial enough to be deemed significant. This could potentially be attributed to the relatively smaller sample size (Dutch = 23 and Indian = 23).

6.2 Relation between Urgency and Dominance

Hypothesis four (H4), which examined the relationship between Dominance and Urgency, revealed a positive correlation between these variables. This relationship was analyzed using Kendall's τ and hypothesis testing, leading to accepting the Hypothesis. The results indicated support for the hypothesis with a strong correlation coefficient indicating that with increase in Urgency of these SFUs corresponds to an increased perception of Dominance, and vice versa. Thus, Dominance can be seen as a behavioural strategy in HRI for conveying Urgency through SFUs, influencing humans to prioritize the robot and thereby facilitating goal achievements when time is critical. This finding aligns with Boos et al.(2020) [101], who discussed dominance as a justifiable behaviour for urgency in task settings that is also valid for sounds.

The close association between vocal expressions of emotions (such as anger, fear, disgust, happiness, sadness, and surprise) and facial expressions, as mentioned by Sauter et al. [117], extends this association to non-verbal auditory interactions like SFUs. This supports the validity of using SFUs to convey emotions effectively.

The higher correlation of these SFUs for Urgency vs. Dominance could have a major contribution from the BEST database as it was expected that NLUs might evoke associations in participants due to their usage in emergency warnings or alerts such as sirens using auditory icons and earcons [13], [20] which are universal sounds and have been present since forever, which are familiar contexts for individuals. And these SFUs from BEST might came across as having similar sounds to it.

Familiarity with such sound patterns and their universality might have added to the higher correlation strength between Dominance and Urgency here. This was further confirmed by the qualitative feedback received from the participants in Table 5.4 where participants from both the cultural backgrounds of Indian and Dutch found the sound to be like an ”alarm”, ”emergency siren”, ”highly urgent” that conveys Urgency.

6.3 Effect of Supporting Measurement Scales

The additional urgency scale, Urgency_Scale2, adopted from Orthmann et al.'s study [2], was used to validate our findings on urgency perception and the intent communication of the robot. Consistently higher ratings (Section 5.3) further support the results from Urgency_Scale1, indicating that the SFUs conveying anger from the EMOGIB and BEST corpora effectively communicated urgency and the intent of the hospital delivery robot.

Given that anger is both a reaction to and a component of perceived Rudeness across cultures, as discussed in Section 2.4.3, it was crucial to evaluate this dimension. The findings suggest that Rudeness ratings were generally lower and less variable compared to those for Dominance and Urgency (Urgency_Scale1). The correlation between Urgency and Rudeness, discussed in Section 5.3.2, shows a very weak positive relationship. This suggests that an increase in Urgency does not necessarily lead to an increase in Rudeness, and vice versa. However, this might not be true always. For example, Sound_7, which scores highest on Urgency_Scale1, also ranks high in Rudeness, as illustrated in the combined plot of all scales' mean

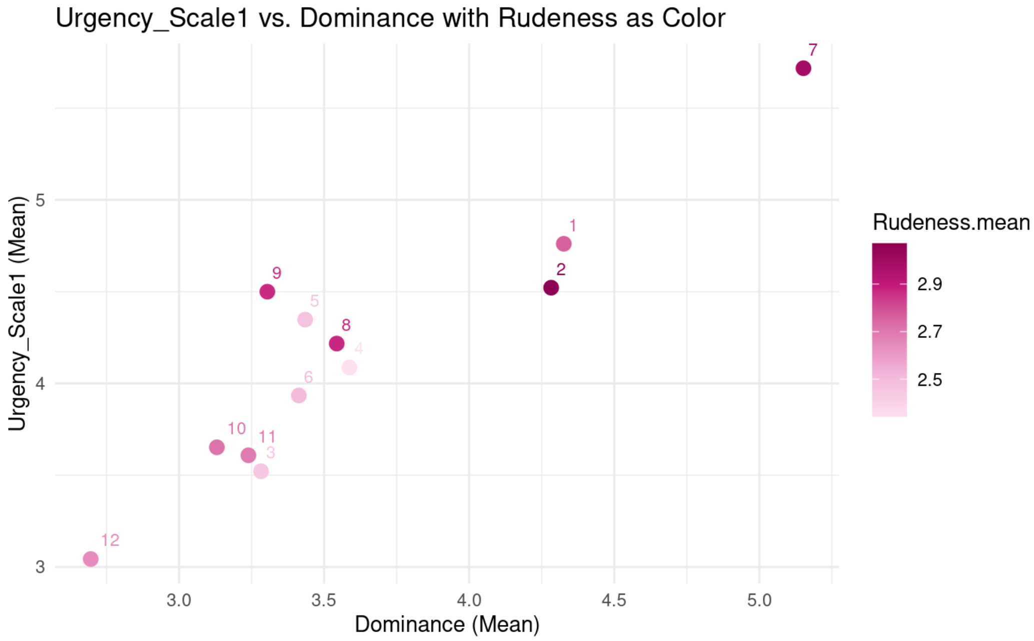


Figure 6.1: Scatter plot representing the relation between Dominance, Urgency_Scale1 and Rudeness

ratings in Figure 6.1. In contrast, some sounds that are moderate to high in Urgency, such as Sound_4, do not score high in Rudeness.

The correlation between Rudeness and Dominance is positive, but Kendall τ indicated that it is low to moderate. This suggests that Dominance can somewhat predict Rudeness, but the relationship is not consistent. As Dominance increases, Rudeness also tends to increase, but not consistently. Thus, it can be said that although Dominant sounds may often be perceived as rude, this is not always the case.

This variability is also reflected in the Thematic analysis in Table 5.4, where some participants described the tone of voice of EMOGIB as rude or shouting. In contrast, others found it polite and empathetic. These findings imply that although Dominance can influence perceptions of Rudeness, however, other factors also play a significant role like tone, pitch, and pace.

Chapter 7

Conclusion, Limitation and Recommendations

This study explored the utilization of Semantic-Free Utterances SFUs by a hospital delivery robot to convey urgency across two cultural groups, Dutch and Indian. The primary research question addressed was:

"What is the effect of the type of Dominant Semantic-Free Utterances from two different corpora (BEST vs. EMOGIB) on individuals from two distinct cultural backgrounds (Indian vs. Dutch) when these utterances are utilized by a hospital delivery robot?"

SRQ_1 *(Database Focused): What is the effect of dominant SFUs from two different corpora (EMOGIB vs. BEST) on the perception of urgency in robot behaviour?*
SRQ_2 *(Culture Focused): What is the effect of dominant SFUs in robot behaviour on the perception of urgency across Dutch and Indian cultural backgrounds?*

The findings highlight the potential of SFUs as a universal auditory communication tool for robots in hospital settings (H-HRI) and other multicultural settings. By incorporating SFUs, social robots can communicate intentions more effectively without the need of direct human-robot dialogue. It ensures inclusiveness with diverse user populations, ultimately enhancing human-robot collaboration.

This study additionally demonstrated the relationship between emotions and intent communication, which should be taken into account when selecting or designing SFUs for various cultural and situational circumstances.

7.1 Implications in HRI

7.1.1 Design Recommendations and Guidelines for Communicating Urgency

The findings from 6.3 are crucial because Dominance and associated behaviours, such as anger, could negatively affect the social acceptability of robots by making them sound unpleasant, rude, or harsh. Dominance and Anger are also highly positively correlated (refer to 4.3) as well as the strong correlation between Dominance and Urgency perception offers practical guidance for designing intent communication.

The results indicate that when designing SFUs or choosing them for conveying Urgency, the intensity of Anger and/or Dominance should be kept moderate to low, to sound Urgent but low on Rudeness. Lower Anger and/or Dominance in robot communication can enhance user trust and comfort, making interactions more pleasant and acceptable. This can be adjusted by controlling parameters such as tone of voice, pitch and pace will result in a sound with a desirable level of Urgency as mentioned by participants in the thematic analysis in Table 5.4. Incorporating user feedback into the design process can help fine-tune Urgent communication to be both effective and socially acceptable by being pleasant.

These offer recommendations for developing more effective communication strategies for conveying Urgency, especially for robots in hospital environments, where communication must be effective yet non-disruptive.

7.1.2 Use of Open-Source Corpora

This thesis supports the use of open-source corpora for conveying intent and urgency. This is particularly beneficial with limited time, and limited resource situations for designing sounds or creating corpora for specific use-case needs. Open-source corpora can reduce costs and accelerate the deployment of robots in real-world applications.

7.1.3 Potential Benefits of SFUs in Healthcare Settings

This study highlights that SFUs, which have been evaluated in child-human robot interaction (cHRI) by Zaga et al. [86], could be particularly beneficial in healthcare environments. Hospitals serving children and patients with neurological disorders such as dementia or autism could especially benefit from the use of SFUs, as these utterances depend on one's cognitive abilities to interpret and decode (Section 2.3.3).

Furthermore, employing SFUs could help alleviate 'alarm fatigue' [1] among medical staff by preventing overlap with more urgent sounds, such as emergency alarms or ambulance sirens. This is crucial in hospital settings, where distinguishing between different levels of Urgency ensures that critical alerts are not masked by the delivery robot's communication, contributing to a more efficient and manageable auditory environment.

7.1.4 Potential of SFUs in Multicultural Settings

The findings underscore the potential of SFUs as a universal auditory communication tool for service robots in multicultural environments, such as hospitals (H-HRI) and public spaces. By overcoming language barriers, SFUs allow robots to convey their intentions inclusively across various cultures, facilitating effective collaboration with diverse groups.

7.2 Limitations

This study provided valuable insights into the influence of cultural backgrounds on urgency perception, but it also had some limitations. For instance, the low R-squared value in the analysis in Section 5.3.1 indicates a possibility that other variables may play a significant role in urgency perception that were not used in the model, indicating that the factors influencing perception are more complex than what it captured. This could be due to the complexity of culture, and considering sound parameters might be beneficial.

Additionally, the methodological limitation of this study is its reliance on online experiments due to the unavailability of the Harmony robot during the study. These experiments depended on participant's visualization using illustrations of potential scenarios along with audio via online survey rather than direct interactions with the robot in the same physical space [103], limiting its ecological validity [122]. Although online experiments have the advantage of being practical, they however offer less control over the user's environment and devices, and factors like device loudness can affect the perception of Urgency as commented by some participants.

Another limitation lies in the use of self-reported data through online questionnaires, which may not accurately reflect participant's true thoughts [43] regarding Urgency perception. More robust measures, such as behavioural or physiological responses, could provide a more detailed understanding of Urgency perception.

[1] https://www.tudelft.nl/en/stories/articles/a-beautiful-alarm-beside-your-hospital-bed/

7.3 Future Research Directions

This study has highlighted the significance of auditory interaction for conveying Urgency in H-HRI, yet a more holistic approach to HRI could yield more benefits in the future.

The perception of robot communication is influenced by various factors, including the robot's morphology, embodiment, voice [123], environment, prior experiences with similar situations etc. All these factors could impact how urgency is perceived in human-robot interactions. Future studies could investigate *multi-modal communication*, by integrating sound with other modalities such as visual effects, lights, facial expressions (if required), gestures, and even haptic feedback. These combined sensory inputs would provide a more complete understanding of user perception and a more sophisticated emotional engagement with robots. This is especially crucial when also considering people with hearing impairments or cognitive challenges, who may benefit from a multi-sensory approach.

This leads to another key area for future research, i.e., *accessibility and inclusiveness*. Investigating the efficacy of accessible and inclusive design practices by conducting user studies with sensitive users using language-independent SFUs in HRI, could confirm their effectiveness for diverse user groups. This would include addressing users with cognitive impairments and those from different age groups, such as older adults (65 and above) and younger users (under 18), especially in sensitive environments like hospitals and childcare settings (cHRI).

Another avenue for future research could be the development of standardized guidelines for sound design in HRI. These guidelines would assist robot designers in crafting appropriate auditory cues. Additionally, a sound design workbook for robots in HRI could aid developers in selecting suitable sounds for robots as they become more ubiquitous in the near future.

In-person testing in realistic environments or close enough to the settings where the robot will be deployed could be incorporated into future studies, particularly in sensitive settings like hospitals. Conducting studies in such settings and considering factors such as noise, lighting, and user stress levels, can lead to more accurate findings and more effective robot behaviour designs and hence leading to efficient HRI.

Conducting longitudinal studies could also provide deeper insights into the long-term effects of SFUs on user perception and behaviour. By observing how users adapt to and interact with robots over extended periods, researchers can better understand how urgency cues evolve in effectiveness over time. As according to Triandis's framework [55] and research on Dutch society suggests that individualism tends to increase with age, although, in collectivism, it tends to decrease [56]. So, longitudinal studies along with considering age as a factor could be considered.

This approach could also reveal how cultural factors, such as those influenced by globalization, shape and transform user responses to these cues over time. Understanding these dynamics would aid in designing robots that are more adaptable, culturally sensitive, and effective across diverse and changing populations.

To conclude, this study offers valuable insights into the role of SFUs in conveying urgency across two cultural backgrounds. By understanding and leveraging these insights, future research can enhance HRI and H-HRI, ultimately leading to the development of more effective, intuitive, and ethically sound service robots capable of operating in diverse, multicultural, and sensitive settings.

Use of Tool

Grammarly has been utilized as a plugin to check the grammar and sentence corrections and enhance the overall writing.

Bibliography

[1] “Enhancing healthcare with assistive robotic mobile manipulation — harmony project — fact sheet — h2020 — cordis — european commission,” https://cordis.europa.eu/project/id/101017008, 2021, accessed: May 12, 2023.

[2] B. Orthmann, I. Leite, R. Bresin, and I. Torre, “Sounding robots: Design and evaluation of auditory displays for unintentional human-robot interaction,” *J. Hum.-Robot Interact.*, aug 2023, just Accepted. [Online]. Available: https://doi.org/10.1145/3611655

[3] A. Weiss and V. Evers, “Exploring cultural factors in human-robot interaction: A matter of personality,” in *International Workshop on Comparative Informatics (IWCI-2011), Copenhagen, Denmark*, vol. 700, 2011.

[4] G.-Z. Yang, J. Bellingham, P. E. Dupont, P. Fischer, L. Floridi, R. Full, N. Jacobstein, V. Kumar, M. McNutt, R. Merrifield *et al.*, “The grand challenges of science robotics,” *Science robotics*, vol. 3, no. 14, p. eaar7650, 2018.

[5] M. Salem, M. Ziadee, and M. Sakr, “Marhaba, how may i help you? effects of politeness and culture on robot acceptance and anthropomorphization,” *ACM/IEEE International Conference on Human-Robot Interaction*, pp. 74–81, 2014.

[6] K. S. Haring, D. Silvera-Tawil, T. Takahashi, M. Velonaki, and K. Watanabe, “Perception of a humanoid robot: A cross-cultural comparison,” *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, vol. 2015-November, pp. 821–826, 11 2015. [Online]. Available: https://dl.acm.org/doi/10.1109/ROMAN.2015.7333613

[7] H. R. Lee and S. Sabanović, “Culturally variable preferences for robot design and use in south korea, turkey, and the united states,” in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’14. New York, NY, USA: Association for Computing Machinery, 2014, p. 17–24. [Online]. Available: https://doi.org/10.1145/2559636.2559676

[8] V. Evers, H. C. Maldonado, T. L. Brodecki, and P. J. Hinds, "Relational vs. group self-construal: Untangling the role of national culture in hri," *HRI 2008 - Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction: Living with Robots*, pp. 255–262, 2008. [Online]. Available: https://dl.acm.org/doi/10.1145/1349822.1349856

[9] G. Hofstede, G. J. Hofstede, and M. Minkov, *Cultures and organizations: Software of the mind*. Mcgraw-hill New York, 2005, vol. 2.

[10] V. Lim, M. Rooksby, and E. S. Cross, "Social robots on a global stage: Establishing a role for culture during human–robot interaction," *International Journal of Social Robotics*, vol. 13, pp. 1307–1333, 9 2021.

[11] M. A. Goodrich and A. C. Schultz, "Human-robot interaction: A survey," *Foundations and Trends in Human-Computer Interaction*, vol. 1, pp. 203–275, 2007.

[12] F. Robinson, H. Pelikan, K. Watanabe, L. Damiano, O. Bown, and M. Velonaki, "Introduction to the special issue on sound in human-robot interaction," *J. Hum.-Robot Interact.*, vol. 12, no. 4, dec 2023. [Online]. Available: https://doi.org/10.1145/3632185

[13] S. Yilmazyildiz, R. Read, T. Belpeame, and W. Verhelst, "Review of semantic-free utterances in social human‚Äìrobot interaction," *International Journal of Human‚ÄìComputer Interaction*, vol. 32, no. 1, pp. 63–85, 2016. [Online]. Available: https://doi.org/10.1080/10447318.2015.1093856

[14] A. KHOTA, A. KIMURA, and E. COOPER, "Modelling of non-linguistic utterances for machine to human communication in dialogue," *International Symposium on Affective Science and Engineering*, vol. ISASE2019, pp. 1–4, 2019.

[15] R. Read and T. Belpaeme, "People interpret robotic non-linguistic utterances categorically," *International Journal of Social Robotics*, vol. 8, pp. 31–50, 1 2016. [Online]. Available: https://link.springer.com/article/10.1007/s12369-015-0304-0

[16] S. Yilmazyildiz, D. Henderickx, B. Vanderborght, W. Verhelst, E. Soetens, and D. Lefeber, "Emogib: Emotional gibberish speech database for affective human-robot interaction," in *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction - Volume Part II*, ser. ACII'11. Berlin, Heidelberg: Springer-Verlag, 2011, p. 163–172.

[17] N. Savela, T. Turja, R. Latikka, and A. Oksanen, “Media effects on the perceptions of robots,” *Human Behavior and Emerging Technologies*, vol. 3, pp. 989–1003, 12 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/hbe2.296https://onlinelibrary.wiley.com/doi/abs/10.1002/hbe2.296https://onlinelibrary.wiley.com/doi/10.1002/hbe2.296

[18] A. B. Latupeirissa and R. Bresin, “Understanding non-verbal sound of humanoid robots in films,” 2020. [Online]. Available: https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-278689

[19] A. G. C. Gonzalez, W. S. Lo, and I. Mizuuchi, “Talk to kotaro: a web crowdsourcing study on the impact of phone and prosody choice for synthesized speech on human impression,” *RO-MAN 2022 - 31st IEEE International Conference on Robot and Human Interactive Communication: Social, Asocial, and Antisocial Robots*, pp. 244–251, 2022.

[20] M. A. Goodrich and A. C. Schultz, “Human‚Äìrobot interaction: A survey,” *Foundations and Trends¬Æ in Human‚ÄìComputer Interaction*, vol. 1, no. 3, pp. 203–275, 2008. [Online]. Available: http://dx.doi.org/10.1561/1100000005

[21] R. Sanz-Segura, E. M. Pérez, and E. Özcan, “Alarm compliance in healthcare: Design considerations for actionable alarms (in intensive care units),” *Proceedings of the Design Society: International Conference on Engineering Design*, vol. 1, pp. 839–846, 2019. [Online]. Available: https://www.cambridge.org/core/journals/proceedings-of-the-international-conference-on-engineering-design/article/alarm-compliance-in-healthcare-design-considerations-for-actionable-alarms-in-intensive DE80475FDB14E8599CF11E13EC34E834

[22] E. Oñzcan, D. Birdja, and J. R. Edworthy, “A holistic and collaborative approach to audible alarm design,” *Biomedical Instrumentation & Technology*, vol. 52, no. 6, pp. 422–432, 2018. [Online]. Available: https://array.aami.org/doi/abs/10.2345/0899-8205-52.6.422

[23] D. Oyserman and S. W. Lee, “Does culture influence what and how we think? effects of priming individualism and collectivism,” *Psychological Bulletin*, vol. 134, pp. 311–342, 3 2008.

[24] S. Hall, *Cultural studies 1983: A theoretical history*. Duke University Press, 2016.

[25] M. Kyrarini, F. Lygerakis, A. Rajavenkatanarayanan, C. Sevastopoulos, H. R. Nambiappan, K. K. Chaitanya, A. R. Babu, J. Mathew, and F. Makedon, "A survey of robots in healthcare," *Technologies*, vol. 9, p. 8, 1 2021.

[26] J. Holland, L. Kingston, C. McCarthy, E. Armstrong, P. O'dwyer, F. Merz, and M. McConnell, "Service robots in the healthcare sector," *Robotics 2021, Vol. 10, Page 47*, vol. 10, p. 47, 3 2021. [Online]. Available: https://www.mdpi.com/2218-6581/10/1/47

[27] V. Ciupe and I. Maniu, "New trends in service robotics," *Mechanisms and Machine Science*, vol. 16, pp. 57–74, 2014. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-01592-7_5

[28] D. Robotics, "Moxi — diligent robotics," https://www.diligentrobots.com/moxi, Accessed on 2023-03-07.

[29] M. Alseddiqi, B. AlMannaei, O. Najam, and A. Al-Mofleh, "The importance of medical robots in improving healthcare services," in *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*, Dec 2022, pp. 1397–1401.

[30] ABB, "Abb demonstrates concept of mobile laboratory robot for hospital of the future," https://new.abb.com/news/detail/37279/hospital-of-the-future, Accessed on 2023-03-07.

[31] D. L. Johanson, H. S. Ahn, and E. Broadbent, "Improving interactions with healthcare robots: A review of communication behaviours in social and healthcare contexts," *International Journal of Social Robotics 2020 13:8*, vol. 13, pp. 1835–1850, 11 2020. [Online]. Available: https://link.springer.com/article/10.1007/s12369-020-00719-9

[32] M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. te Boekhorst, and K. L. Koay, "Avoiding the uncanny valley: Robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion," *Autonomous Robots*, vol. 24, pp. 159–178, 11 2008. [Online]. Available: https://link.springer.com/article/10.1007/s10514-007-9058-3

[33] S. R. Fussell, S. Kiesler, L. D. Setlock, and V. Yew, "How people anthropomorphize robots," in *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, 2008, pp. 145–152.

[34] B. R. Duffy, "Anthropomorphism and the social robot," *Robotics and Autonomous Systems*, vol. 42, pp. 177–190, 3 2003.

[35] C. Breazeal, “Toward sociable robots,” *Robotics and Autonomous Systems*, vol. 42, pp. 167–175, 3 2003.

[36] K. A. Barchard, L. Lapping-Carr, R. S. Westfall, A. Fink-Armold, S. B. Banisetty, and D. Feil-Seifer, “Measuring the perceived social intelligence of robots,” *J. Hum.-Robot Interact.*, vol. 9, no. 4, sep 2020. [Online]. Available: https://doi.org/10.1145/3415139

[37] R. Stock-Homburg, “Survey of emotions in human–robot interactions: Perspectives from robotic psychology on 20 years of research,” *International Journal of Social Robotics 2021 14:2*, vol. 14, pp. 389–411, 6 2021. [Online]. Available: https://link.springer.com/article/10.1007/s12369-021-00778-6

[38] E. Hildt, “What sort of robots do we want to interact with? reflecting on the human side of human-artificial intelligence interaction,” *Frontiers in Computer Science*, vol. 3, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcomp.2021.671012

[39] E. Wiese, G. Metta, and A. Wykowska, “Robots as intentional agents: using neuroscientific methods to make robots appear more social,” *Frontiers in psychology*, vol. 8, p. 281017, 2017.

[40] C. Mcginn and I. Torre, “Can you tell the robot by the voice? an exploratory study on the role of voice in the perception of robots,” pp. 211–221. [Online]. Available: http://www.diva-portal.orghttp://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-270674

[41] E. Pacherie and V. F. Castro, “Robots and resentment: Commitments, recognition and social motivation in hri.” [Online]. Available: https://hal.science/ijn_03496738

[42] S. Šabanović, “Robots in society, society in robots: Mutual shaping of society and technology as a framework for social robot design,” *International Journal of Social Robotics*, vol. 2, pp. 439–450, 2010.

[43] C. . Diana, F. Kawahara, M. Saccardi, and I. E. Kret, “A cross-cultural comparison on implicit and explicit attitudes towards artificial agents,” *International Journal of Social Robotics*, 2022. [Online]. Available: https://doi.org/10.1007/s12369-022-00917-7

[44] C. Bartneck, T. Suzuki, T. Kanda, and T. Nomura, “The influence of people’s culture and prior experiences with aibo on their attitude towards robots,” *AI and Society*, vol. 21, pp. 217–230, 1 2007. [Online]. Available: https://link.springer.com/article/10.1007/s00146-006-0052-7

[45] M. Lantis, “The silent language,” 1959.

[46] A. Boeijen, M. Sonneveld, and C. Hao, “Culture sensitive design education: the best of all worlds,” 09 2017.

[47] S. Ervin-Tripp, F. Kluckhohn, and F. Strodtbeck, “Variations in value orientation,” *The American Journal of Psychology*, vol. 76, p. 342, 06 1963.

[48] N. York, C. San, F. Lisbon, L. Madrid, M. City, M. New, D. San, J. S. Singapore, S. Toronto, G. Hofstede, G. J. Hofstede, and M. Minkov, “Cultures and organizations software ofthe mind intercultural cooperation and its importance for survival —mc igrauu jhill,” 2010.

[49] G. Hofstede, *Culture’s consequences: International differences in work-related values*. sage, 1984, vol. 5.

[50] G. Hofstede and M. H. Bond, “Hofstede’s culture dimensions: An independent validation using rokeach’s value survey,” *Journal of Cross-Cultural Psychology*, vol. 15, pp. 417–433, 1984.

[51] R. Cohen, *Negotiating across cultures: International communication in an interdependent world*. US institute of Peace Press, 1997.

[52] P. B. yuen Ngai, “Nonverbal communication behavior of professional administrators from ethiopia tanzania hong kong and china in negotiations with u.s. negotiators: Cross-cultural perspectives,” 1996. [Online]. Available: https://api.semanticscholar.org/CorpusID:157006807

[53] G. Hofstede, *Culture’s consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage publications, 2001.

[54] C. Liu, “Chinese,why don‚Äôt you show your anger? ‚Äî a comparative study between chinese and americans in expressing anger,” *International Journal of Social Science and Humanity*, vol. 4, pp. 206–209, 01 2014.

[55] H. C. Triandis *et al.*, “Culture and social behavior,” 1994.

[56] L. Oppenheimer, “Perception of individualism and collectivism in dutch society: A developmental approach,” *International Journal of Behavioral Development*, vol. 28, no. 4, pp. 336–346, 2004. [Online]. Available: https://doi.org/10.1080/01650250444000009

[57] S. Hareli, K. Kafetsios, and U. Hess, “A cross-cultural study on emotion expression and the learning of social norms,” *Frontiers in Psychology*, vol. 6, p. 152022, 10 2015.

[58] R. Gogineni, R. Kallivayalil, S. Sharma, S. Rataemane, and S. Akhtar, "Globalization of culture: Impact on indian psyche," *Indian Journal of Social Psychiatry*, vol. 34, p. 303, 2018. [Online]. Available: https://journals.lww.com/ijsp/fulltext/2018/34040/globalization_of_culture__impact_on_indian_psyche.13.aspx

[59] M. Javidan, R. J. House, P. W. Dorfman, P. J. Hanges, and M. S. D. Luque, "Conceptualizing and measuring cultures and their consequences: a comparative review of globe's and hofstede's approaches," *Journal of International Business Studies*, vol. 37, pp. 897–914, 2006. [Online]. Available: http://www.thunderbird.edu/wwwfiles/

[60] X. Fang, M. Rychlowska, and J. Lange, "Cross-cultural and inter-group research on emotion perception," *Journal of Cultural Cognitive Science*, vol. 6, no. 1, pp. 1–7, 2022.

[61] S. Andrist, M. Ziadee, H. Boukaram, B. Mutlu, and M. Sakr, "Effects of culture on the credibility of robot speech: A comparison between english and arabic english site." [Online]. Available: http://dx.doi.org/10.1145/2696454.2696464

[62] C. Darwin, *The Expression of the Emotions in Man and Animals. With Photographic and Other Illustrations. London, J. Murray, 1872*. Culture et civilisation, 1969.

[63] P. Ekman, "Cross-cultural studies of facial expression: A century of research in review academic press," *New York, NY*, 1973.

[64] C. E. Izard, "The face of emotion." 1971.

[65] S. Zhang and M. D. Pell, "Cultural differences in vocal expression analysis: Effects of task, language, and stimulus-related factors," *PLOS ONE*, vol. 17, p. e0275915, 10 2022. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0275915

[66] E. Sanoubari and J. E. Young, "Explicit, neutral, or implicit: A cross-cultural exploration of communication-style preferences in human robot interaction," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 237–238. [Online]. Available: https://doi.org/10.1145/3173386.3177061

[67] I. KieÇbasiewicz-Drozdowska and S. I. Radko, "The role of intercultural communication in tourism and recreation." *Studies in physical Culture*

&Tourism, vol. 13, pp. 75–85, 2006. [Online]. Available: https://api.semanticscholar.org/CorpusID:54055843

[68] D. Matsumoto, "Cultural influences on the perception of emotion," *http://dx.doi.org/10.1177/0022022189201006*, vol. 20, pp. 92–105, 3 1989. [Online]. Available: https://journals.sagepub.com/doi/10.1177/0022022189201006

[69] I. Fernández, P. Carrera, F. Sánchez, D. Paez, and L. Candia, "Differences between cultures in emotional verbal and non-verbal reactions 1," *Psicothema*, vol. 12, pp. 83–92, 2000.

[70] R. Merkin, "The relationship between individualism / collectivism consultation and harmony needs," *Journal of Intercultural Communication*, vol. 2015, 11 2015.

[71] S. Nishimura, A. Nevgi, and S. Tella, "Communication style and cultural features in high/low context communication cultures: A case study of finland, japan and india," 01 2008.

[72] K. R. Scherer, R. Banse, and H. G. Wallbott, "Emotion inferences from vocal expression correlate across languages and cultures," *Journal of Cross-cultural psychology*, vol. 32, no. 1, pp. 76–92, 2001.

[73] M. D. Pell, L. Monetta, S. Paulmann, and S. A. Kotz, "Recognizing emotions in a foreign language," *Journal of Nonverbal Behavior*, vol. 33, pp. 107–120, 2009.

[74] G. Cosme, V. Tavares, G. Nobre, C. Lima, R. Sá, P. Rosa, and D. Prata, "Cultural differences in vocal emotion recognition: a behavioural and skin conductance study in portugal and guinea-bissau," *Psychological Research*, vol. 86, no. 2, pp. 597–616, 2022.

[75] P. Laukka, H. Elfenbein, N. Söder, H. Nordström, J. Althoff, W. Chui, F. Iraki, T. Rockstuhl, and N. Thingujam, "Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations," *Frontiers in psychology*, vol. 4, p. 353, 07 2013.

[76] M. Combi, "Cultures and technology: An analysis of some of the changes in progress-digital, global and local culture," *Cultural Heritage in a Changing World*, pp. 3–15, 1 2016.

[77] C. C. Bennett, Y. H. Bae, J. H. Yoon, Y. Chae, E. Yoon, S. Lee, U. Ryu, S. Y. Kim, and B. Weiss, "Effects of cross-cultural language differences on social

cognition during human-agent interaction in cooperative game environments," *Computer Speech and Language*, vol. 81, 6 2023.

[78] S. Šabanović, C. C. Bennett, and H. R. Lee, "Towards culturally robust robots: A critical social perspective on robotics and culture."

[79] T. Ramge, "Who's afraid of ai? : fear and promise in the age of thinking machines." [Online]. Available: https://books.google.com/books/about/Who_s_Afraid_of_AI.html?id=OFtsDwAAQBAJ

[80] B. Bramas, Y. M. Kim, and D. S. Kwon, "Design of a sound system to increase emotional expression impact in human-robot interaction," *2008 International Conference on Control, Automation and Systems, ICCAS 2008*, pp. 2732–2737, 2008.

[81] M. Schwenk and K. O. Arras, "R2-d2 reloaded: A flexible sound synthesis system for sonic human-robot interaction design," *IEEE RO-MAN 2014 - 23rd IEEE International Symposium on Robot and Human Interactive Communication: Human-Robot Co-Existence: Adaptive Interfaces and Systems for Daily Life, Therapy, Assistance and Socially Engaging Interactions*, pp. 161–167, 10 2014.

[82] F. A. Robinson, O. Bown, and M. Velonaki, "Implicit communication through distributed sound design: Exploring a new modality in human-robot interaction," *ACM/IEEE International Conference on Human-Robot Interaction*, pp. 597–599, 3 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/3371382.3377431

[83] T. Hermann, A. Hunt, J. G. Neuhoff *et al.*, *The sonification handbook*. Logos Verlag Berlin, 2011, vol. 1.

[84] B. Schuller and A. Batliner, *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing*, 1st ed. Wiley Publishing, 2013.

[85] S. Yilmazyildiz, L. Latacz, W. Mattheyses, and W. Verhelst, "Expressive gibberish speech synthesis for affective human-computer interaction," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6231 LNAI, pp. 584–590, 2010.

[86] C. Zaga, R. A. D. Vries, S. J. Spenkelink, K. P. Truong, and V. Evers, "Help-giving robot behaviors in child-robot games: Exploring semantic free utter-

ances," *ACM/IEEE International Conference on Human-Robot Interaction*, vol. 2016-April, pp. 541–542, 4 2016.

[87] X. Liu, J. Dong, and M. Jeon, "Robots' "woohoo" and "argh" can enhance users' emotional and social perceptions: An exploratory study on non-lexical vocalizations and non-linguistic sounds," *ACM Transactions on Human-Robot Interaction*, vol. 12, pp. 1–20, 12 2023. [Online]. Available: https://doi.org/10.1145/3626185

[88] H. R. Pelikan, M. Broth, and L. Keevallik, ""are you sad, cozmo?" how humans make sense of a home robot's emotion displays," *ACM/IEEE International Conference on Human-Robot Interaction*, pp. 461–470, 3 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/3319502.3374814

[89] B. J. Zhang and N. T. Fitter, "Nonverbal sound in human-robot interaction: A systematic review," *J. Hum.-Robot Interact.*, feb 2023, just Accepted. [Online]. Available: https://doi.org/10.1145/3583743

[90] C. Breazeal, *Designing Sociable Robots*. Cambridge, MA, USA: MIT Press, 2002.

[91] S. A. Brewster, "Using non-speech sound to overcome information overload," *Displays*, vol. 17, pp. 179–189, 5 1997.

[92] T. Komatsu, K. Kobayashi, S. Yamada, K. Funakoshi, and M. Nakano, "Can users live with overconfident or unconfident systems? a comparison of artificial subtle expressions with human-like expression," in *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 1595–1600. [Online]. Available: https://doi.org/10.1145/2212776.2223678

[93] A. De Angeli, G. Johnson, and L. Coventry, "The unfriendly user: Exploring social reactions to chatterbots," 11 2001.

[94] J. S. A. Lee, M. F. B. Abbas, C. K. Seow, Q. Cao, K. P. Yar, S. L. Keoh, and I. Mcloughlin, "Non-verbal auditory aspects of human-service robot interaction," pp. 11–12, 2021. [Online]. Available: https://doi.org/10.1109/SOLI54607.2021.9672366.https://eprints.gla.ac.uk/260481/

[95] T. Tunanunkul, T. Kaizuka, R. Zheng, and K. Nakano, "The influence of audio warning urgency and situational urgency on collision avoidance performance," *International Journal of Automotive Engineering*, vol. 9, pp. 165–172, 09 2018.

[96] E. Hellier and J. Edworthy, “On using psychophysical techniques to achieve urgency mapping in auditory warnings,” *Applied Ergonomics*, vol. 30, pp. 167–171, 1999.

[97] J. Edworthy, “The design and implementation of non-verbal auditory warnings,” *Applied Ergonomics*, vol. 25, pp. 202–210, 1994.

[98] E. J. Hellier, J. Edworthy, and I. Dennis, “Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency,” *Human Factors*, vol. 35, no. 4, pp. 693–706, 1993, pMID: 8163282. [Online]. Available: https://doi.org/10.1177/001872089303500408

[99] S. J. Isherwood and D. McKeown, “Semantic congruency of auditory warnings,” *Ergonomics*, vol. 60, pp. 1014–1023, 7 2017. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/00140139.2016.1237677

[100] U.S. Department of Health and Human Services. (2017, January) Cultural and linguistic competence: Communication styles. Accessed on 06-10-2023. [Online]. Available: https://thinkculturalhealth.hhs.gov/assets/pdfs/resource-library/communication-styles.pdf

[101] A. Boos, M. Sax, and J. Reinhardt, “Investigating perceived task urgency as justification for dominant robot behaviour,” in *HCI International 2020 - Posters*, C. Stephanidis and M. Antona, Eds. Cham: Springer International Publishing, 2020, pp. 117–124.

[102] J. P. Dillard, D. H. Solomon, and M. T. Palmer, “Structuring the concept of relational communication,” *Communications Monographs*, vol. 66, no. 1, pp. 49–65, 1999.

[103] J. Li, W. Ju, and C. Nass, “Observer perception of dominance and mirroring behavior in human-robot relationships,” *ACM/IEEE International Conference on Human-Robot Interaction*, vol. 2015-March, pp. 133–140, 3 2015.

[104] J. Li, A. Cuadra, B. Mok, B. Reeves, J. Kaye, and W. Ju, “Communicating dominance in a nonanthropomorphic robot using locomotion,” *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 8, 3 2019. [Online]. Available: https://dl.acm.org/doi/10.1145/3310357

[105] J. K. Burgoon, M. L. Johnson, and P. T. Koch, “The nature and measurement of interpersonal dominance,” *Communications Monographs*, vol. 65, no. 4, pp. 308–335, 1998.

[106] C. Nass, Y. Moon, and N. Green, "Are machines gender neutral? gender-stereotypic responses to computers with voices," *Journal of applied social psychology*, vol. 27, no. 10, pp. 864–876, 1997.

[107] B. Mutlu and J. Forlizzi, "Robots in organizations: The role of workflow, social, and environmental factors in human-robot interaction," *HRI 2008 - Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction: Living with Robots*, pp. 287–294, 2008. [Online]. Available: https://dl.acm.org/doi/10.1145/1349822.1349860

[108] J. V. Doorn, M. Zeelenberg, and S. M. Breugelmans, "Anger and prosocial behavior," *http://dx.doi.org/10.1177/1754073914523794*, vol. 6, pp. 261–268, 3 2014. [Online]. Available: https://journals.sagepub.com/doi/10.1177/1754073914523794

[109] H. C. Lench, N. T. Reed, T. George, K. A. Kaiser, and S. G. North, "Anger has benefits for attaining goals." *Journal of Personality and Social Psychology*, 10 2023.

[110] J. C. C. Cabral and R. M. M. de Almeida, "Effects of anger on dominance-seeking and aggressive behaviors," *Evolution and Human Behavior*, vol. 40, pp. 23–33, 1 2019.

[111] J. C. C. Cabral, P. de Souza Tavares, and R. M. M. de Almeida, "Reciprocal effects between dominance and anger: A systematic review," *Neuroscience and biobehavioral reviews*, vol. 71, pp. 761–771, 12 2016. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27984056/

[112] M. Boiger, S. De Deyne, and B. Mesquita, "Emotions in ‚Äúthe world‚Äù: cultural practices, products, and meanings of anger and shame in two individualist cultures," *Frontiers in Psychology*, vol. 4, 2013. [Online]. Available: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2013.00867

[113] C. Günsoy, S. E. Cross, A. K. Uskul, G. Adams, and B. Gercek-Swing, "Avoid or fight back? cultural differences in responses to conflict and the role of collectivism, honor, and enemy perception," *Journal of Cross-Cultural Psychology*, vol. 46, no. 8, pp. 1081–1102, 2015. [Online]. Available: https://doi.org/10.1177/0022022115594252

[114] C. Günsoy, "Rude bosses versus rude subordinates: How we respond to them depends on our cultural background," *International Journal of Conflict Management*, vol. 31, pp. 175–199, 3 2020.

[115] T. Y. Kim and D. L. Shapiro, "Retaliation against supervisory mistreatment: Negative emotion, group membership, and cross-cultural difference," *International Journal of Conflict Management*, vol. 19, pp. 339–358, 2008.

[116] N. Hashimoto, E. Hagens, A. Zgonnikov, and M. L. Lupetti, "Safe spot: Perceived safety of dominant and submissive appearances of quadruped robots in human-robot interactions," 2024.

[117] D. A. Sauter, F. Eisner, A. J. Calder, and S. K. Scott, "Perceptual cues in nonverbal vocal expressions of emotion," *Quarterly journal of experimental psychology*, vol. 63, no. 11, pp. 2251–2272, 2010.

[118] S. Serholt, W. Barendregt, T. Ribeiro, G. Castellano, A. Paiva, A. Kappas, R. Aylett, and F. Nabais, "Emote: Emboided-perceptive tutors for empathy-based learning in game environment," in *European Conference GBL, Porto, Portugal*, 2013.

[119] L. G. Fox and S. D. Dalebout, "Use of the median method to enhance detection of the mismatch negativity in the responses of individual listeners," *Journal of the American Academy of Audiology*, vol. 13, no. 02, pp. 083–092, 2002.

[120] K. A. Barchard, L. Lapping-Carr, R. S. Westfall, A. Fink-Armold, S. B. Banisetty, and D. Feil-Seifer, "Measuring the perceived social intelligence of robots," *J. Hum.-Robot Interact.*, vol. 9, no. 4, sep 2020. [Online]. Available: https://doi.org/10.1145/3415139

[121] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only anova procedures," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 143–146. [Online]. Available: https://doi.org/10.1145/1978942.1978963

[122] R. Hartson and P. S. Pyla, "Chapter 14 - rigorous empirical evaluation: Preparation," in *The UX Book*, R. Hartson and P. S. Pyla, Eds. Boston: Morgan Kaufmann, 2012, pp. 503–536. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123852410000142

[123] R. G. Read and T. Belpaeme, "Interpreting non-linguistic utterances by robots: Studying the influence of physical appearance," *AFFINE'10 - Proceedings of the 3rd ACM Workshop on Affective Interaction in Natural Environments, Co-located with ACM Multimedia 2010*, pp. 65–70, 2010.

Appendix A: Pre-Study

Several screenshots of the Pre-Study survey conducted in Qualtrics are shared.

Welcome!, Dear Participant,

About the survey:
This survey is part of a master's thesis in Interaction Technology at the University of Twente, which focuses on finding the most suitable semantic-free utterances to convey urgency for a hospital delivery robot. In this pre-study, we focus on identifying which sounds (semantic-free utterances) can convey anger and dominance for a hospital delivery robot.
Semantic-free utterances (SFUs) represent a form of auditory interaction or communication that can convey emotion and intent expression, composed of vocalizations and sounds without semantic content. In simple terms, semantic-free utterances (SFUs) refer to sounds that express emotions or intentions without using specific words in any language. Because they are independent of language semantics, semantic-free utterances hold the potential to serve as an alternative to natural language for machines, especially robots, particularly in cross-cultural environments.. In this study, two types of SFUs from two databases are used: Emote, which contains non-linguistic utterances (NLU), and EMOGIB, which contains gibberish speech (GS).

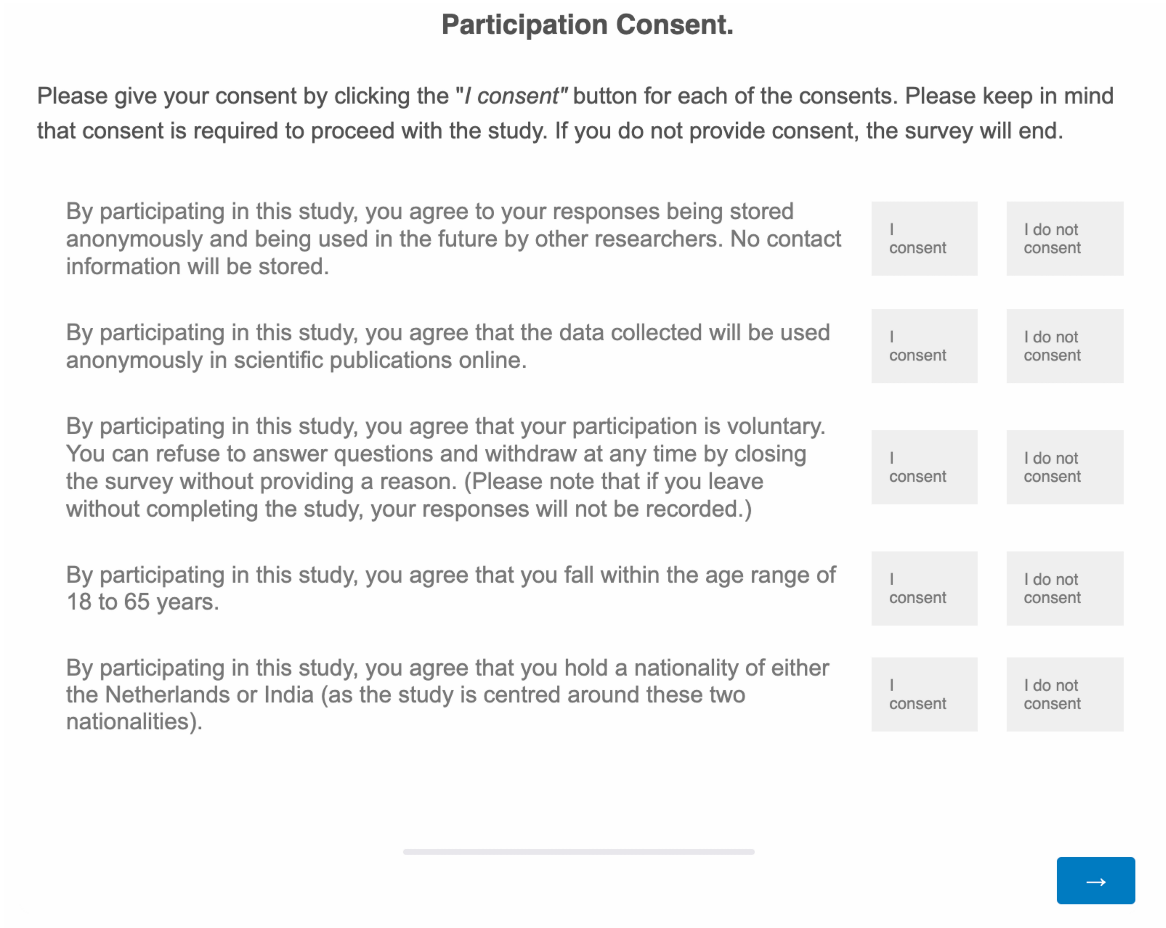
Join us in exploring the intersection of human-robot interaction, culture, and emotions by offering your valuable feedback. Please read the instructions carefully before starting the survey.

Instructions for the participants: This study is for individuals between the ages of *18 and 65* who are nationals of either the *Netherlands or India.* No personal information that can be used to identify an individual is collected in the study. The data will be anonymized and saved for future research. Participation is voluntary, and you can opt out of the survey at any point in time by simply closing the browser. Incomplete surveys will not be recorded. Once the survey is completed, your data will be saved and cannot be withdrawn. This survey will take ~30 minutes of your time. There will be 32 sounds. Make sure you are listening to them via a good-quality headset or speaker. To proceed with the survey, you must digitally sign online consent by answering a few questions anonymously. Thank you very much for volunteering and giving your precious time to this survey. I hope you enjoy the sound.

Contact Information:
In case of any query, feel free to contact me via email:
Ishitaa Narwane (Master's Student): i.narwane@student.utwente.nl
or my supervisors: dr. Khiet Truong (HMI), *University of Twente* (k.p.truong@utwente.nl) and Hideki Garcia Goo (HMI), *University of Twente* (h.garciagoo@utwente.nl) If you have questions about your rights as a research participant or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee Information & Computer Science at ethicscommittee-cis@utwente.nl.

UNIVERSITY OF TWENTE.

Figure A.1: Introduction Page of the Pre-study Qualtrics Survey



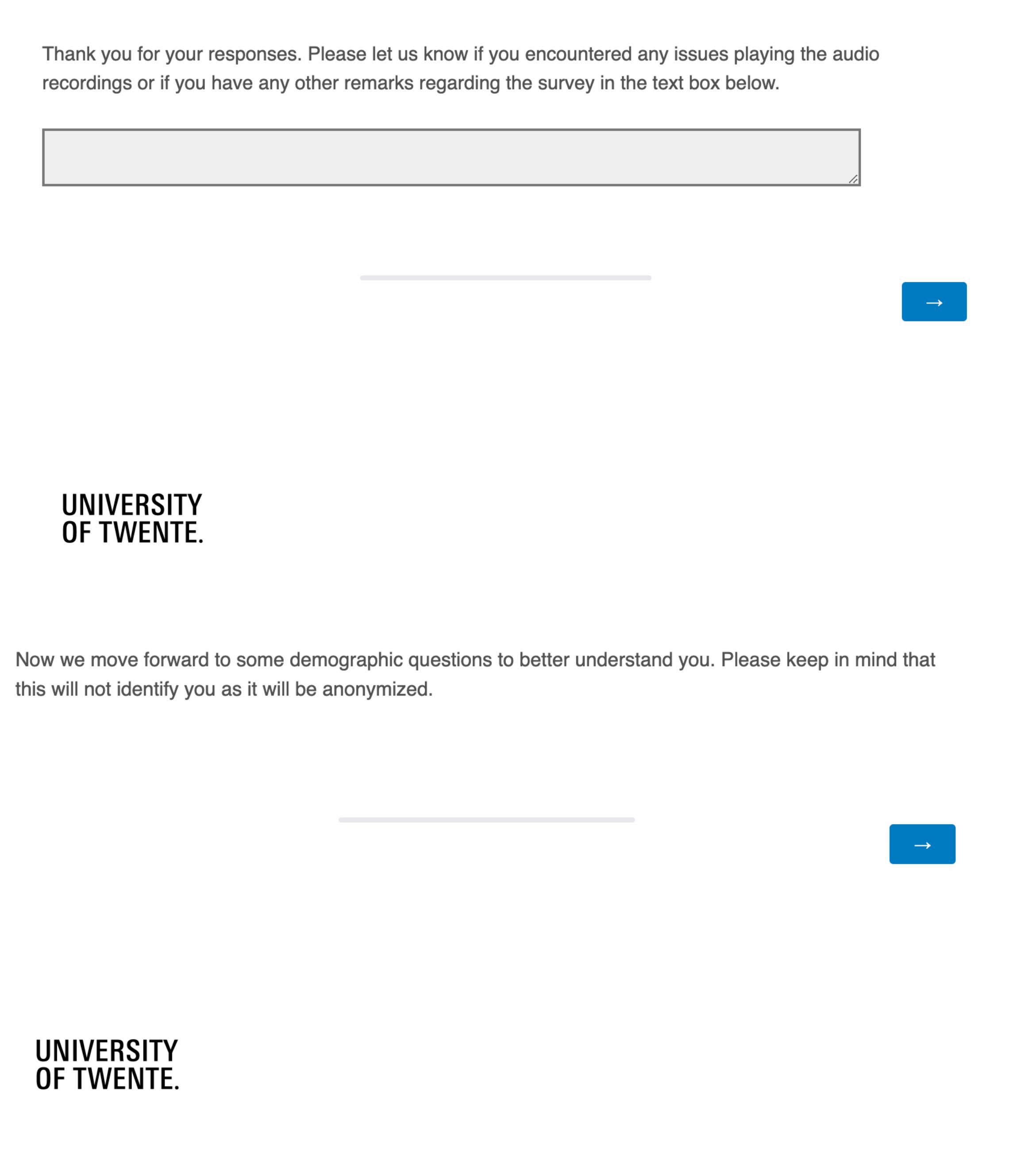
Participation Consent.

Please give your consent by clicking the "*I consent*" button for each of the consents. Please keep in mind that consent is required to proceed with the study. If you do not provide consent, the survey will end.

By participating in this study, you agree to your responses being stored anonymously and being used in the future by other researchers. No contact information will be stored.	I consent	I do not consent
By participating in this study, you agree that the data collected will be used anonymously in scientific publications online.	I consent	I do not consent
By participating in this study, you agree that your participation is voluntary. You can refuse to answer questions and withdraw at any time by closing the survey without providing a reason. (Please note that if you leave without completing the study, your responses will not be recorded.)	I consent	I do not consent
By participating in this study, you agree that you fall within the age range of 18 to 65 years.	I consent	I do not consent
By participating in this study, you agree that you hold a nationality of either the Netherlands or India (as the study is centred around these two nationalities).	I consent	I do not consent

→

Figure A.2: Online Consent of Pre-study



Thank you for your responses. Please let us know if you encountered any issues playing the audio recordings or if you have any other remarks regarding the survey in the text box below.

→

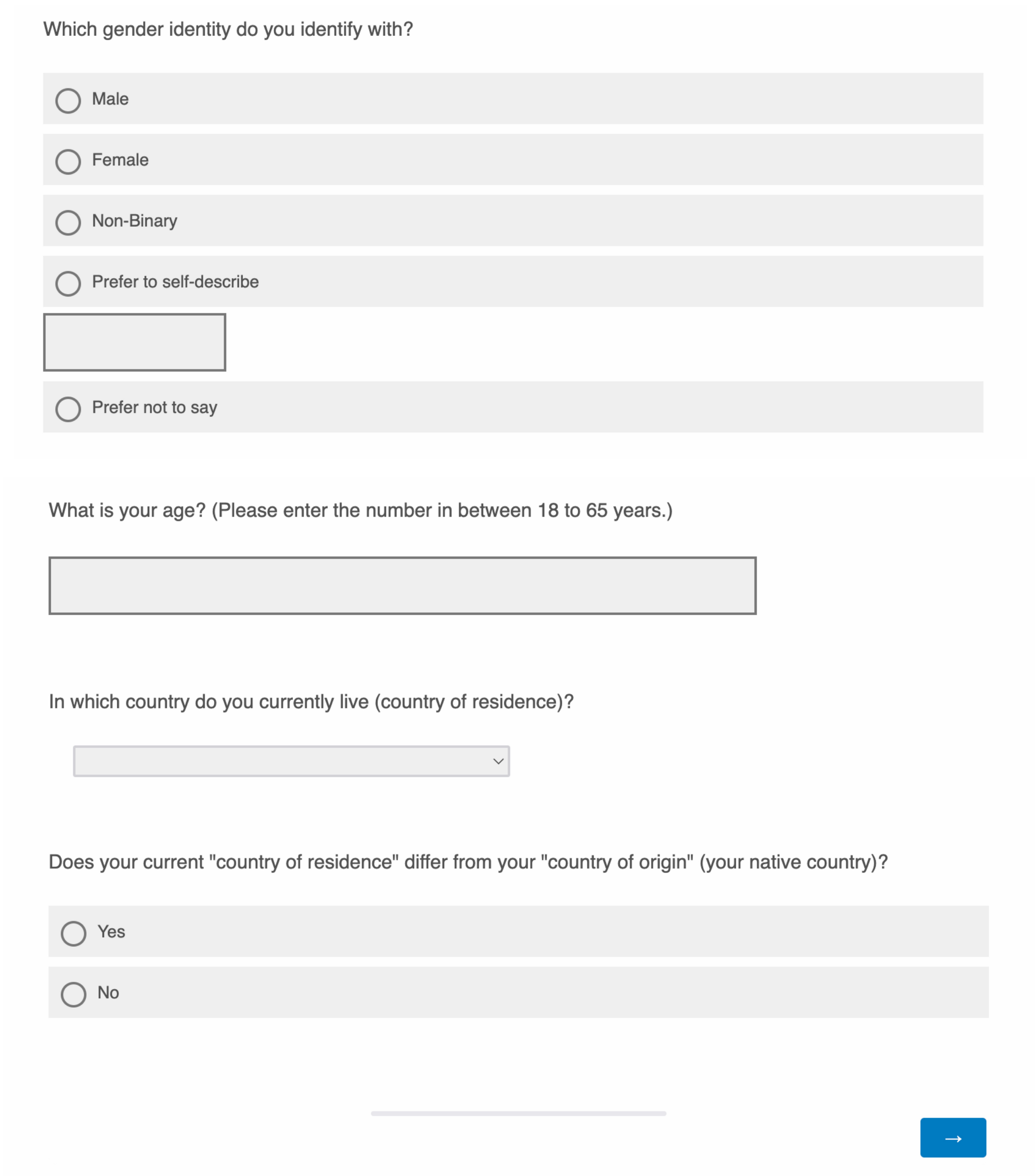
UNIVERSITY OF TWENTE.

Now we move forward to some demographic questions to better understand you. Please keep in mind that this will not identify you as it will be anonymized.

→

UNIVERSITY OF TWENTE.

Figure A.3: End of Sound Ratings Section in Pre-study


Which gender identity do you identify with?

- Male
- Female
- Non-Binary
- Prefer to self-describe
- Prefer not to say

What is your age? (Please enter the number in between 18 to 65 years.)

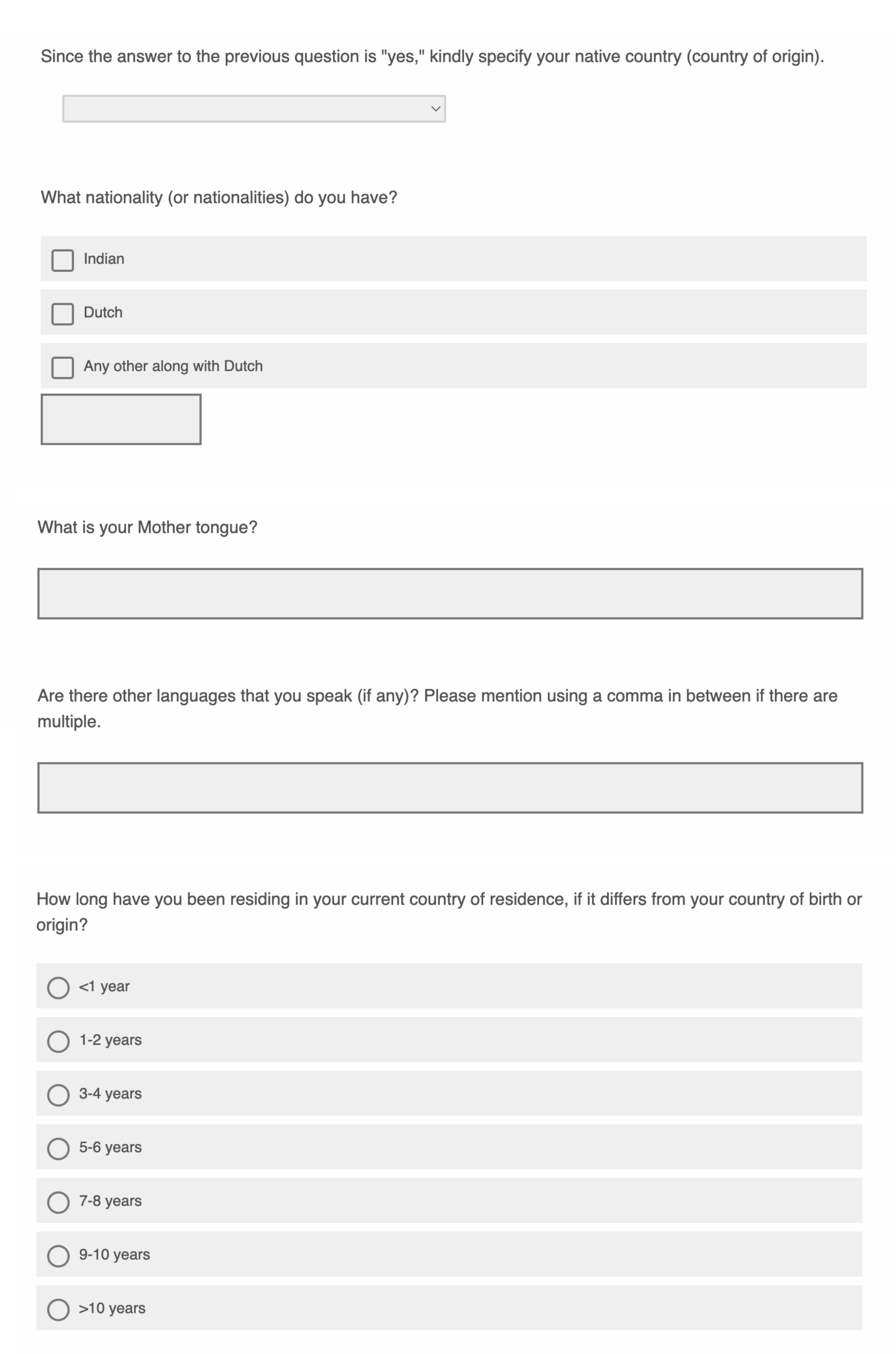
In which country do you currently live (country of residence)?

Does your current "country of residence" differ from your "country of origin" (your native country)?

- Yes
- No

→

Figure A.4: Demographic survey part 1 of Pre-study



Since the answer to the previous question is "yes," kindly specify your native country (country of origin).

What nationality (or nationalities) do you have?

- Indian
- Dutch
- Any other along with Dutch

What is your Mother tongue?

Are there other languages that you speak (if any)? Please mention using a comma in between if there are multiple.

How long have you been residing in your current country of residence, if it differs from your country of birth or origin?

- <1 year
- 1-2 years
- 3-4 years
- 5-6 years
- 7-8 years
- 9-10 years
- >10 years

Figure A.5: Demographic survey part 2 of Pre-study

How long have you resided in your country of origin, if it differs from your current country of residence?

- <1 years
- 1-2 years
- 3-4 years
- 5-6 years
- 7-8 years
- 9-10 years
- >10 years

How familiar are you with robots? (Rate on a scale of 1 to 5)

Not at all Familiar 1	2	3	4	Very familiar 5
○	○	○	○	○

Where does your familiarity with robots originate from? Please type your answer below. (e.g., owning a robot at home, working with robots in a lab, involvement in robot development, watching movies, etc.)

→

Figure A.6: Demographic survey part 3 of Pre-study

De-Briefing

Thank you for taking the time to participate and for taking the time to complete the survey.

This study consisted of two phases: the pre-study and the main study. In the pre-study phase, the sounds were selected randomly from hundreds of sound from the two databases BEST/Emote and EMOGIB. There we examined how sounds could convey anger and dominance, finding that the sounds that conveyed anger also conveyed dominance. The top eight high, medium and low rated dominant semantic-free utterances from the pre-study are selected for communicating urgency and whether that sound would cause humans to move aside and give way to the robot.

This is done to better understand the relationship between urgency, dominance, and anger perceptions of a sound, as well as the influence of cultural factors on these perceptions, by examining two cultures: Indian and Dutch. This main study used two categories of semantic-free utterances: gibberish speech and non-linguistic utterances sourced from the Emote/BEST and EMOGIB databases that can communicate urgency for a hospital delivery robot.

Your participation in this study has contributed valuable insights to our research efforts. Thank you once again for your time and contributions. In case of any query or question, feel free to contact me via email.

Principal researcher:
Ishitaa Narwane (Master's Student): i.narwane@student.utwente.nl
or my supervisors:

1. **dr. Khiet Truong (HMI), *University of Twente,* k.p.truong@utwente.nl and**
2. **Hideki Garcia Goo (HMI), *University of Twente,* h.garciagoo@utwente.nl**

If you have questions about your rights as a research participant or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee Information & Computer Science: ethicscommittee-cis@utwente.nl.

Please note that this is your last chance to withdraw from the survey. If you want to withdraw, please do so by closing the browser. Please select the sentence below for confirmation. Once you have selected the confirmation and clicked submit, your responses will be recorded, and you can no longer delete them.

○ By selecting this statement, I acknowledge and confirm my participation in this survey, and I understand that I cannot withdraw after that.

Submit

Figure A.7: Pre-study De-briefing

Appendix B: Main-Study

Table B.1: Means for Dominance, Urgency, and Rudeness for Each Sound

Sound ID	Database	Dominance Mean	Urgency_Scale1 Mean	Urgency_Scale2 Mean	Rudeness Mean
10	Best	3.130	3.652	3.717	2.717
11	Best	3.239	3.609	3.826	2.696
12	Best	2.696	3.043	3.435	2.647
7	Best	5.152	5.717	5.500	2.995
8	Best	3.543	4.217	4.109	2.875
9	Best	3.304	4.500	4.696	2.870
1	Emogib	4.326	4.761	5.087	2.761
2	Emogib	4.283	4.522	5.326	3.071
3	Emogib	3.283	3.522	4.022	2.451
4	Emogib	3.587	4.087	4.457	2.342
5	Emogib	3.435	4.348	4.891	2.473
6	Emogib	3.413	3.935	4.152	2.500

Table B.2: Medians for Dominance, Urgency, and Rudeness for Each Sound

Sound ID	Database	Dominance Median	Urgency_Scale1 Median	Urgency_Scale2 Median	Rudeness Median
10	Best	3.0	4.0	4.0	3.00
11	Best	3.0	4.0	4.0	2.75
12	Best	2.0	3.0	3.0	2.75
7	Best	5.0	6.0	6.0	3.00
8	Best	3.5	4.0	4.0	3.00
9	Best	3.0	5.0	5.0	3.00
1	Emogib	4.0	5.0	5.0	2.75
2	Emogib	4.0	5.0	5.0	3.00
3	Emogib	3.0	3.5	3.5	2.50
4	Emogib	3.5	4.0	4.0	2.25
5	Emogib	3.0	5.0	5.0	2.50
6	Emogib	3.0	4.0	4.0	2.50

Table B.3: Shapiro-Wilk Test Results

Culture	Database	p-value
Dutch	Best	.000004
Dutch	Emogib	.0000035
Indian	Best	.000024
Indian	Emogib	.0000005

Table B.4: Levene's Test for Homogeneity of Variance

Variable	Df	F value	Pr(¿F)
Dominance	1	3.772	.053
Urgency_Scale1	1	0.693	.406
Urgency_Scale2	1	0.083	.774
Rudeness	1	0.589	.443

Thematic Analysis code for Robot Familiarity

Code	**Groundedness**
Books	3
Development	5
Robot Professional development	4
Software development	1
Education	11
Academic research	8
Engineering	6
Entertainment	23
Documentaries	1
Games	3
Movies	22
Online media	1
Home Environment	8
Personal assistant	1
Robot Vacuum cleaner	3
Smart-home Robot	4
Movies	1
Robotics involvement	7
Interest in robotics	1
Involvement in research	6
Previous robot interaction	2
Surrounding	2
Working with robots	4

Table B.5: Thematic Analysis code for Robot Familiarity

Design of the Study in Qualtrics

Welcome!
Dear Participant,
Thank you for showing interest in this research. Please read the information below carefully before participating in this survey.

About the research:
This survey is part of a master's thesis in interaction technology at the University of Twente, which focuses on finding the most suitable semantic-free utterances to convey urgency for a hospital delivery robot. In this main study, we want to determine which sounds (semantic-free utterances) are appropriate for conveying urgency for a hospital delivery robot.

Semantic-free utterances (SFU) are vocalizations that convey emotion and intent without using specific words (free of language semantics). SFU, such as those from movie characters like R2D2 or Minions, offer a way for robots to communicate independently of language, with the potential to bridge cultural gaps in vocal communication in human-robot interaction.

Your participation will help determine how acceptable these sounds are in real-life hospital settings and among people of different cultures. This survey will take ~30 minutes of your time.

Participation criteria:

- If you have participated in the pre-study of this research, which involved rating sounds for "anger" and "dominance," please do not participate in this study.
- This study is for individuals between the ages of *18 and 65* whose culture identity is either Dutch or Indian.
- This study involves rating sounds, so make sure you are listening to it either via a good-quality headset or speaker (possibly in a quiet place).

Data collection, usage, and retention period:

- No personal information that can be used to identify an individual is collected in the study.
- The data collected is anonymized and will be saved for future research.
- Incomplete surveys will not be recorded.
- Once the survey is completed, your data will be saved and cannot be withdrawn.

Procedure for withdrawal from this survey:

- Participation in this study is completely voluntary.
- You can withdraw from the survey at any time by simply closing your browser.
- Note that if you fail to select consent at the end of the survey, your records will not be considered.

Figure B.1: Main-study's Welcome Page with Instructions Part 1

Contact information:
In case of any query, feel free to contact me via email:
Ishitaa Narwane (Master's Student): i.narwane@student.utwente.nl
or my supervisors:

1. dr. Khiet Truong (HMI), *University of Twente* (k.p.truong@utwente.nl) and
2. Hideki Garcia Goo (HMI), *University of Twente* (h.garciagoo@utwente.nl)

If you have questions about your rights as a research participant or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee Information & Computer Science at ethicscommittee-cis@utwente.nl.

Join us in exploring the intersection of human-robot interaction, culture, and expression by offering valuable feedback. To proceed with the survey, you must digitally sign the online consent that you will see at the end of this page. If you do not select “I consent," the study will end.

UNIVERSITY OF TWENTE.

Figure B.2: Main-study’s Welcome Page with Instructions Part 2

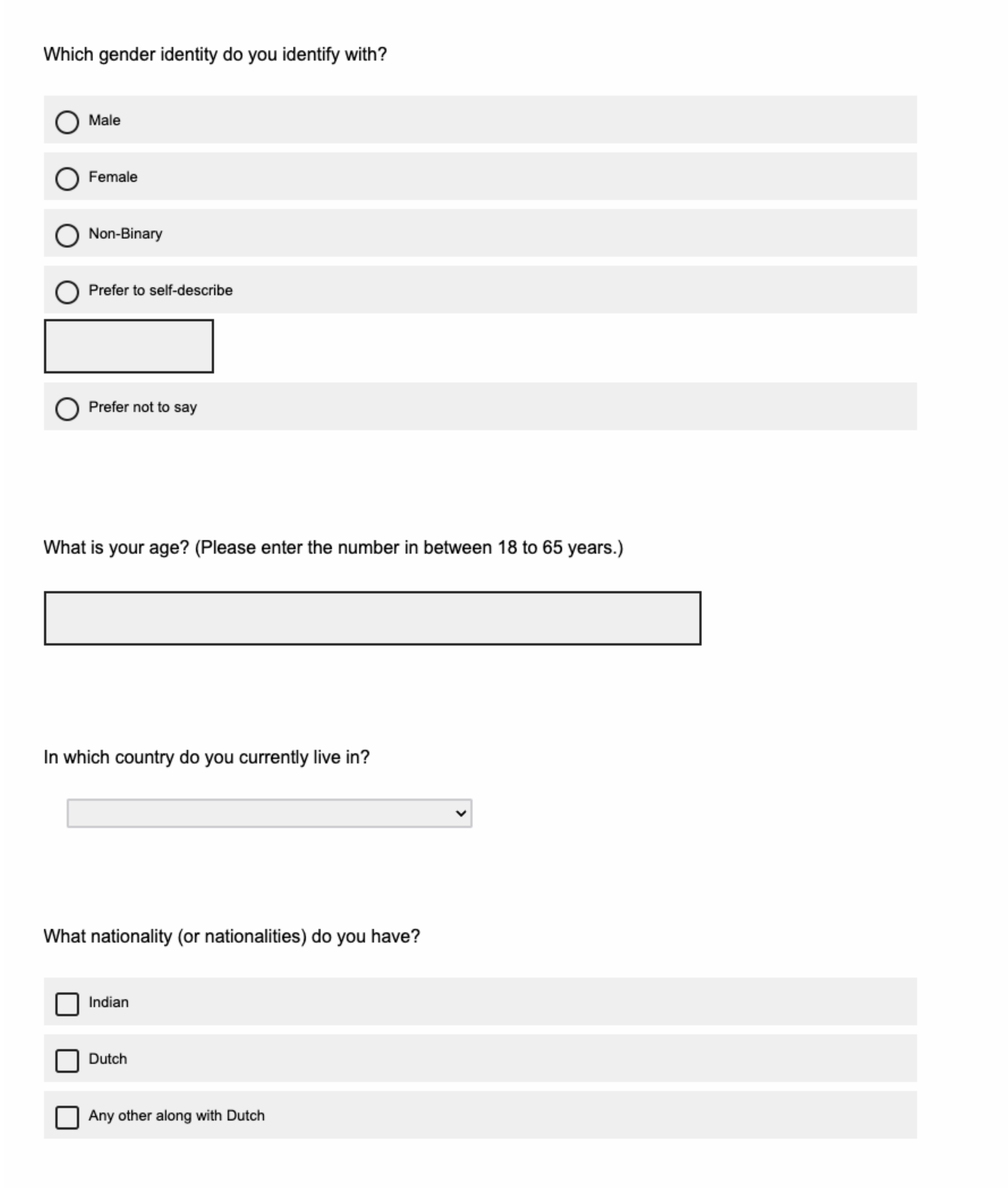
Participation Consent.

Please give your consent by clicking the "*I consent"* button for each of the consents. Please keep in mind that consent is required to proceed with the study. If you do not provide consent, the survey will end.

By participating in this study, you agree that the data collected can be used anonymously in scientific publications.	I consent	I do not consent
By participating in this study, you understand that your participation is voluntary. You can refuse to answer questions and withdraw at any time by closing the survey without providing a reason. (Please note that if you leave without completing the study, your responses will not not be used.)	I consent	I do not consent
By participating in this study, you agree that your feedback can be used as quotes in scientific publications available online.	I consent	I do not consent

→

Figure B.3: Main-study’s Online Consent Form



Which gender identity do you identify with?

- Male
- Female
- Non-Binary
- Prefer to self-describe
- Prefer not to say

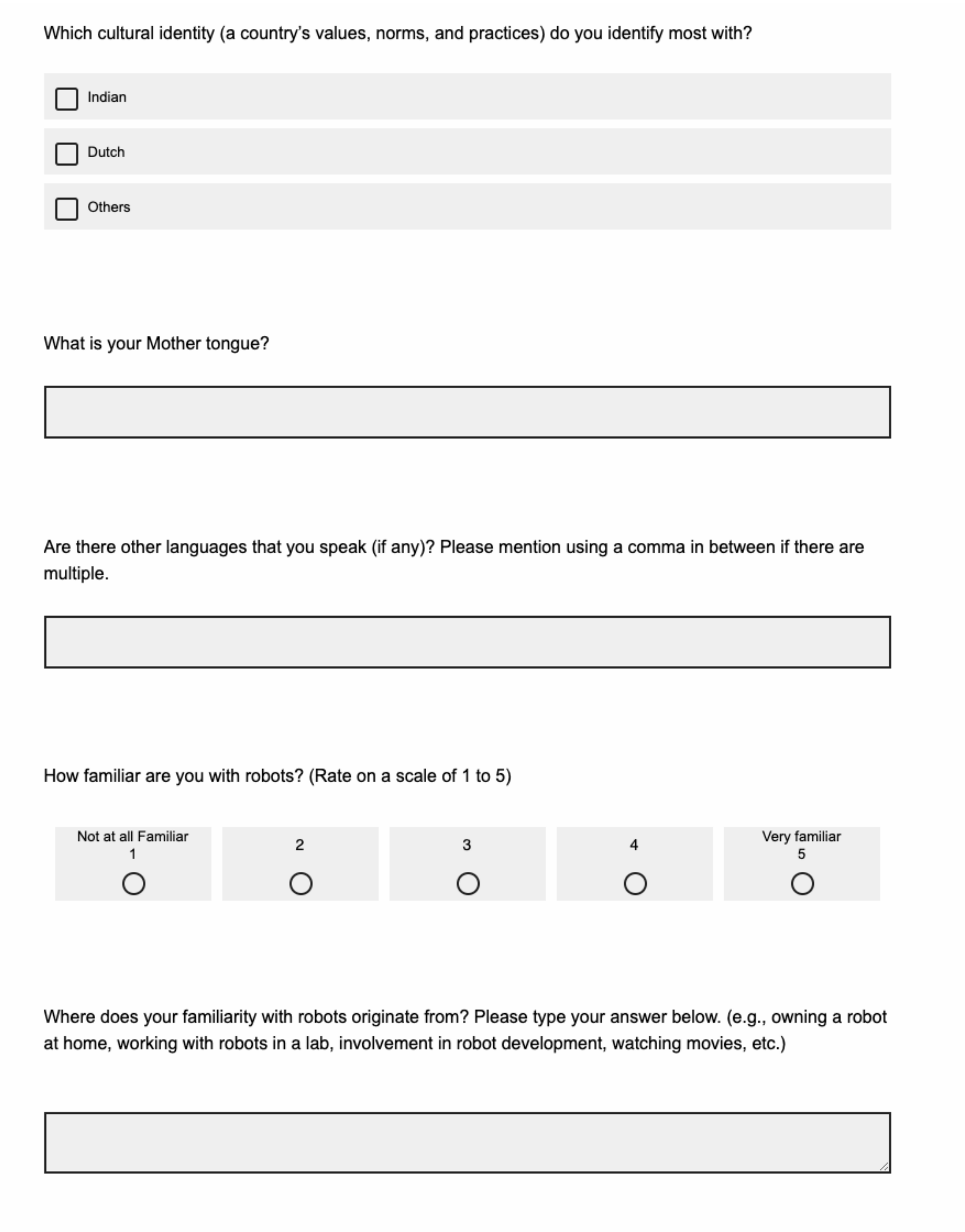
What is your age? (Please enter the number in between 18 to 65 years.)

In which country do you currently live in?

What nationality (or nationalities) do you have?

- Indian
- Dutch
- Any other along with Dutch

Figure B.4: Main-study's Demographic Questionnaire 1


Which cultural identity (a country's values, norms, and practices) do you identify most with?

- [ ] Indian
- [ ] Dutch
- [ ] Others

What is your Mother tongue?

Are there other languages that you speak (if any)? Please mention using a comma in between if there are multiple.

How familiar are you with robots? (Rate on a scale of 1 to 5)

Not at all Familiar 1	2	3	4	Very familiar 5
○	○	○	○	○

Where does your familiarity with robots originate from? Please type your answer below. (e.g., owning a robot at home, working with robots in a lab, involvement in robot development, watching movies, etc.)

Figure B.5: Main-study's Demographic Questionnaire 2

Below, you can see images of the Harmony robot. It is a mobile robot (1.45 meters tall) that supports logistics and laboratory staff in transporting biomedical samples (e.g., test tubes) between laboratories in a hospital.

Please listen to the following sounds:

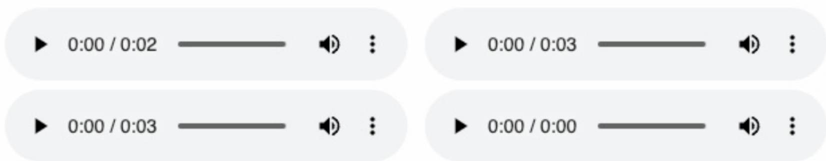
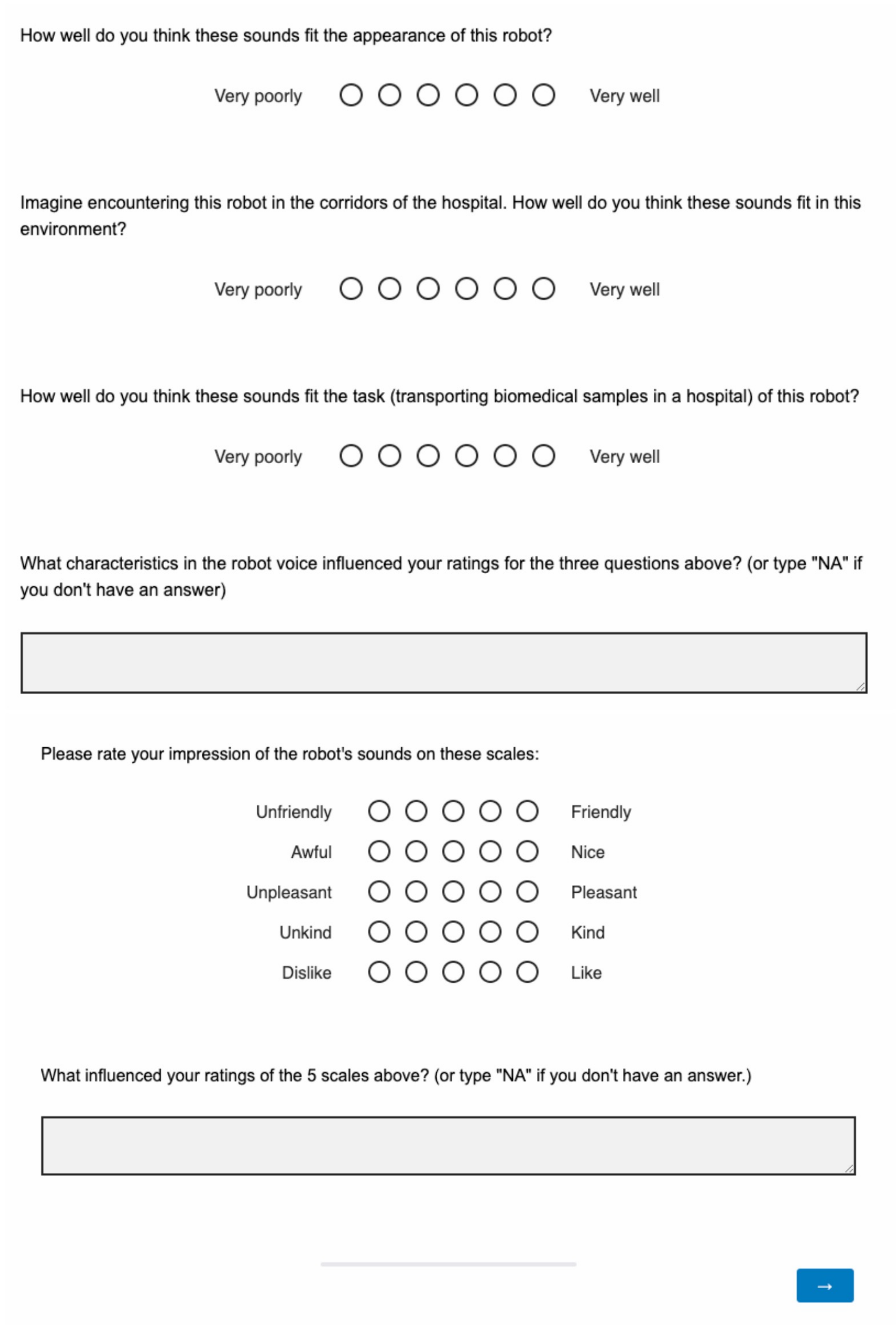


Figure B.6: Other Measurements 1 (Not discussed in this research)


How well do you think these sounds fit the appearance of this robot?

Very poorly ○ ○ ○ ○ ○ ○ Very well

Imagine encountering this robot in the corridors of the hospital. How well do you think these sounds fit in this environment?

Very poorly ○ ○ ○ ○ ○ ○ Very well

How well do you think these sounds fit the task (transporting biomedical samples in a hospital) of this robot?

Very poorly ○ ○ ○ ○ ○ ○ Very well

What characteristics in the robot voice influenced your ratings for the three questions above? (or type "NA" if you don't have an answer)

Please rate your impression of the robot's sounds on these scales:

Unfriendly	○	○	○	○	○	Friendly
Awful	○	○	○	○	○	Nice
Unpleasant	○	○	○	○	○	Pleasant
Unkind	○	○	○	○	○	Kind
Dislike	○	○	○	○	○	Like

What influenced your ratings of the 5 scales above? (or type "NA" if you don't have an answer.)

→

Figure B.7: Other Measurements 2(Not discussed in this research)

De-Briefing

Thank you for taking the time to participate and for taking the time to complete the survey.

This study consisted of two phases: the pre-study and the main study. In the pre-study phase, we examined how sounds could convey anger and dominance, finding that the sounds that conveyed anger also conveyed dominance. The top eight highly rated dominant semantic-free utterances from the pre-study are selected for communication urgency and whether that sound would cause humans to move aside and give way to the robot.

This is done to better understand the relationship between urgency, dominance, and anger perceptions of a sound, as well as the influence of cultural factors on these perceptions, by examining two cultures: Indian and Dutch. This main study used two categories of semantic-free utterances: gibberish speech and non-linguistic utterances sourced from the Emote and EMOGIB databases that can communicate urgency for a hospital delivery robot.

Your participation in this study has contributed valuable insights to our research efforts. Thank you once again for your time and contributions. In case of any query or question, feel free to contact me via email.

Principal researcher:
Ishitaa Narwane (Master's Student): i.narwane@student.utwente.nl
or my supervisors:

1. **dr. Khiet Truong (HMI),** ***University of Twente,*** **k.p.truong@utwente.nl and**
2. **Hideki Garcia Goo (HMI),** ***University of Twente,*** **h.garciagoo@utwente.nl**

If you have questions about your rights as a research participant or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee Information & Computer Science: ethicscommittee-cis@utwente.nl.

Please note that this is your last chance to withdraw from the survey. If you want to withdraw, please do so by closing the browser. Please select the sentence below for confirmation. Once you have selected the confirmation and clicked submit, your responses will be recorded, and you can no longer delete them.

○ By selecting this statement, I acknowledge and confirm my participation in this survey, and I understand that I cannot withdraw after that.

Submit

UNIVERSITY
OF TWENTE.

Figure B.8: Main-study's De-Briefing