Nonverbal Sound in Human-Robot Interaction: a Systematic Review

BRIAN J. ZHANG and NAOMI T. FITTER, Collaborative Robotics and Intelligent Systems (CoRIS) Institute, USA

Nonverbal sound offers great potential to enhance robots' interactions with humans, and a growing body of research has begun to explore nonverbal sound for tasks such as sound source localization, explicit communication, and improving sociability. However, nonverbal sound has a broad interpretation and design space that can draw from areas such as machine learning, music theory, and foley. We sought to identify and compare use cases and approaches for nonverbal sound in human-robot interaction through a systematic review. A search of sound and robotics-related publisher databases yielded 148 peer-reviewed articles presenting systems, studies, and taxonomies. Differences in taxonomy and overlap of terminology with adjacent research fields such as speech, gaze, and gesture posed difficulties for the search, which we attempted to address through a multi-stage search process. Based on the reviewed articles, we developed a pair of taxonomies using scientific communication principles and analyzed study designs and measures for the creation of nonverbal robot sound. We discuss recommendations for the field, including the use of the new taxonomies; methods for design, generation, and validation; and paths for future research. Roboticists may benefit from incorporating nonverbal sound as a key component in multimodal human-robot interaction.

CCS Concepts: • **Human-centered computing → Auditory feedback**; **Sound-based input / output**; • **Computer systems organization → Robotics**.

Additional Key Words and Phrases: nonverbal sound, human-robot interaction, systematic review

ACM Reference Format:
Brian J. Zhang and Naomi T. Fitter. 2022. Nonverbal Sound in Human-Robot Interaction: a Systematic Review. In . ACM, New York, NY, USA, 47 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The sense of hearing provides unique capabilities to humans, including detecting out-of-sight individuals and events, receiving speech and nonverbal sounds, and directing attention toward environmental activity. While each of these abilities has been examined by roboticists, with some research even going beyond human-relevant uses of sound into animal-like echolocation [27], most work on robot sound has focused on speech. In this systematic review, we aimed to examine *nonverbal sound*, defined as audible sound not involving or using words [92, 93], particularly in the realm of human-robot interaction.

Prior threads of research in nonverbal sound for human-robot interaction have drawn inspiration from many adjacent fields, framing, and goals, making it more difficult to unify all nonverbal sound work under a single umbrella. The most direct predecessor to our work reviewed nonverbal sound *explicitly produced* by robots for human-robot interaction [175], a categorization explained further in Section 3.1. Researchers seeking an up-to-date holistic picture of how robots can employ nonverbal sound will benefit from our updated review, and researchers investigating nonverbal sound will benefit from our proposed taxonomies.

In this paper, we seek to answer the questions:

1. How has nonverbal sound been used in human-robot interaction?
2. What barriers exist in nonverbal robot sound research?
3. What are the next steps for nonverbal robot sound research?

We conducted a systematic review to answer these questions, the methods of which are described in Section 2. We begin with analyzing the terminologies and taxonomies used by the works in Section 3.1, which we synthesize into a pair of taxonomies that we used to classify the systematic review results. Section 3.2 details human-centered study methods in the robot sound space and examines overall strengths and weaknesses of past approaches. Section 4 discusses the implications and overall findings of the available research, as well as strengths, weaknesses, and future work for the field.

2 METHODS

A closely related work by Yilmazyildiz et al. in 2016 provided an in-depth but non-systematic review of nonverbal sound, particularly sounds closely related to speech [175]. We sought to provide an updated and broader look at the nonverbal sound space through a systematic review, which we conducted in April 2022. The review methods incorporated terms with awareness of the taxonomy laid out by Yilmazyildiz et al. [175], which is further explained in Section 3.1, as well as recently used terms in the authors' prior works [178, 180] and adjacent fields such as auditory display [52]. Our review followed a four-step process: (1) a keyword search, (2) a relevance scoring process, (3) a manual title and abstract review, (4) a full paper review. The step-by-step review results are provided in the supplementary material included with this paper.

First, we conducted a keyword search for peer-reviewed conference and journal papers using the terms:

```
("human-robot interaction" OR "social robot") AND
(sound* OR soni* OR audi* OR aur* OR acoustic* OR music* OR utter*) AND
(nonverbal OR "non-verbal" OR "non-linguistic")
```

These keywords originated from (1) the context of the question (human-robot interaction); (2) various terms that refer to sounds, such as sonic, auditory, aural, acoustic, music, and utterance; (3) qualifiers to guide the search results toward nonverbal sounds rather than speech.

We searched the Association for Computing Machinery (ACM) Digital Library (663 results), Institute of Electrical and Electronics Engineers (IEEE) Xplore (1608 results), Japan Science and Technology Agency J-STAGE (42 results), SAGE Journals (93 results), Springer Link (1060 results), and Taylor & Francis Online (273 results). As some papers have been published in sound- and music-focused venues rather than robotics, we performed an additional, simpler search with the term "robot" in the proceedings of the International Conference on Auditory Display (55 results) and Sound and Music Computing Conference (5 results) for a total of 3799 results. Abstracts were manually filled in for Springer and Taylor & Francis results. Duplicates and non-articles that could be identified by title or DOI were removed, resulting in 3610 items.

Given the large number of search results, we opted to conduct a replicable relevance scoring process that helped us to identify and more closely review the most pertinent related works. Firstly, as this review focuses on nonverbal sound in the context of robotics, all articles that did not include "robot" in the title or abstract were removed. Next, each article's title and abstract were searched for the number of instances of sound-related keywords used in the keyword search ("sound," "soni," "audi," "aur," "acoustic," "music, " and "utter"), the count of which formed the positive component of the score, and the number of instances of adjacent interaction modalities that would indicate a focus on a topic other than nonverbal sound ("gaz[e]," "gestur[e]," "speech, " "speak," "fac[e]," and "voice"), the count of which formed the negative component of the score. The positive and negative components were summed, and the 349 articles with scores of 0 or greater passed to the next step of the review process.

The remaining articles' titles and abstracts were manually reviewed and sorted by a research assistant for relevance, resulting in 127 "yes," 52 "maybe," and 170 "no" categorizations. Most "no" categorizations were papers that were focused solely on verbal sound upon closer review. "Maybe" articles did not clearly indicate interaction modalities in the title and abstract; these articles were reviewed in their entirety to check for apparent relevance. Of the 52 "maybe" articles, 25 were found to be relevant, receiving an updated "confirmed maybe" categorization. The 127 "yes" and 25 "confirmed maybe" articles were read meticulously in full. After this closest review, a further 4 articles from the "yes" category were excluded after closer review; two were found not to contain content on nonverbal sound in the full text, one contained content but only from a separately included article, and one was not in English. Thus, the final set of papers considered throughout the remainder of this review includes 148 articles.

3 RESULTS

The 148 articles considered in this review are organized chronologically in Table 1. Based on the nonverbal sound topics of the reviewed articles, we developed new taxonomies of sound form and function and categorized each article as described in Section 3.1. We extracted information on study methods for human perceptions of nonverbal robot sounds in Section 3.2 as an extension of the study method review in [175].

Nonverbal sound in human-robot interaction is a young field that continues to grow over time, as seen in Figure 1. Starting from 1996, the annual publication rate has generally increased with a peak in 2016 of 16 articles. This growth is primarily driven by an increase in the number of articles on sound creation. Research on robots concurrently using sound creation and perception, particularly for music, has also begun to appear in the literature in the last decade.

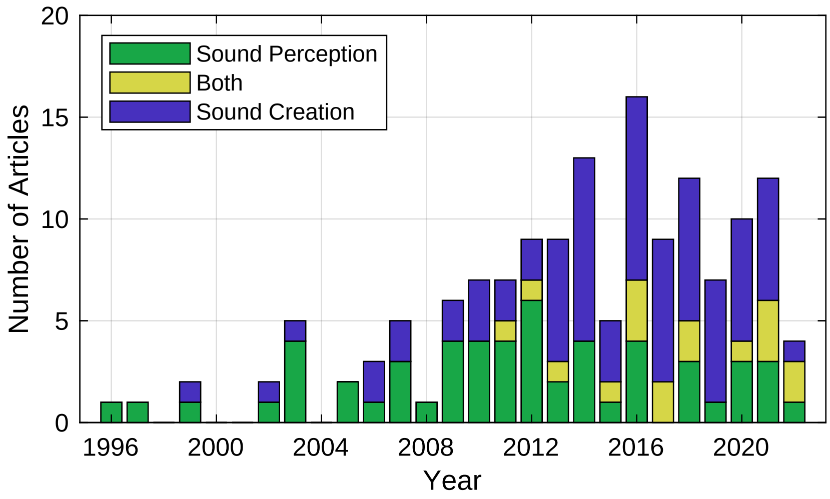

Fig. 1. Number of articles published in each year from 1996 to 2022. Articles are separated by research topic into categories of sound perception, sound creation, and both.

Year	Authors	Title	Function	Form & Techniques
1996	T. Shibata et al. [136]	Emotional robot for intelligent system-artificial emotional creature project	Sound source localization	
1997	T. Shibata et al. [137]	Artificial emotional creature for human-machine interaction	Sound source localization	
1999	R. A. Brooks et al. [15]	The Cog Project: Building a Humanoid Robot	Sound source localization	
1999	A. Camurri et al. [18]	EyesWeb-toward gesture and affect recognition in dance/music interactive systems	Functional robot sound	Music; sonification; personalization
2002	G. Johannsen [66]	Auditory display of directions and states for mobile systems	Functional robot sound	Music; sonification; auditory display
2002	H. G. Okuno et al. [107]	Social Interaction of Humanoid Robot Based on Audio-Visual Tracking	Sound source localization	
2003	T. Hermann et al. [53]	Interactive visualization and sonification for monitoring complex processes	Functional robot sound	Music; artificial sounds; natural sounds; sonification
2003	H. G. Okuno and K. Nakadai [109]	Realizing personality in audio-visually triggered non-verbal behaviors	Sound source localization Sound source separation	
2003	H. G. Okuno et al. [108]	Design and Implementation of Personality of Humanoids in Human Humanoid Non-verbal Interaction	Sound source localization	
2003	H. G. Okuno et al. [105]	Human–robot non-verbal interaction empowered by real-time auditory and visual multiple-talker tracking	Sound source localization	
2003	H. G. Okuno et al. [104]	Real-time Sound Source Localization and Separation based on Active Audio-Visual Integration	Sound source localization Sound source separation	

Year	Authors	Title	Function	Form & Techniques
2005	M. Bennewitz et al. [7]	Towards a humanoid museum guide robot that interacts with multiple persons	Sound source localization	
2005	L. Błażejewski [17]	Spatial Sound Localization for Humanoid	Sound source localization Sound source separation	
2006	D. Brock and E. Martinson [13]	Exploring the utility of giving robots auditory perspective-taking abilities	Sound perception (loudness)	
2006	J. F. Gorostiza et al. [44]	Multimodal Human-Robot Interaction Framework for a Personal Robot	Sound creation (unspecified)	Music
2006	S. Yamada and T. Komatsu [171]	Designing simple and effective expression of robot's primitive minds to a human	Emotional robot sound	Electronic sounds
2007	E. C. Haas [46]	Integrating Auditory Warnings with Tactile Cues in Multimodal Displays for Challenging Environments	Functional robot sound	
2007	K. Kobayashi et al. [71]	Action Sloping as a Way for Users to Notice a Robot's Function	Functional robot sound	Electronic sounds
2007	M. P. Michalowski et al. [95]	A Dancing Robot for Rhythmic Social Interaction	Music recognition (for dance)	
2007	V. M. Trifa et al. [152]	Real-time acoustic source Sound source localization in noisy environments for human-robot multimodal interaction	Sound source localization	
2007	M. Yamamoto and T. Watanabe [173]	Analysis by Synthesis of an Information Presentation Method of Embodied Agent Based on the Time Lag Effects of Utterance to Communicative Actions	Paralanguage recognition	
2008	N. A. Mirza et al. [97]	Developing social action capabilities in a humanoid robot using an interaction history architecture	Sound perception (loudness)	

Year	**Authors**	**Title**	**Function**	**Form & Techniques**
2009	H. D. Kim et al. [68]	Human Tracking System Integrating Sound and Face Sound source localization Using an Expectation-Maximization Algorithm in Real Environments	Sound source localization	
2009	M. P. Michalowski et al. [96]	Rhythmic attention in child-robot dance play	Music recognition (for dance) Music recognition (for games)	
2009	B. Robins et al. [122]	From Isolation to Communication: A Case Study Evaluation of Robot Assisted Play for Children with Autism with a Minimally Expressive Humanoid Robot	Music synthesis (physical)	Instrumental music
2009	J. Solis et al. [141]	Implementation of an Auditory Feedback Control System on an Anthropomorphic Flutist Robot Inspired on the Performance of a Professional Flutist	Music synthesis (physical)	Instrumental music
2009	A. Tapus [147]	Improving the Quality of Life of People with Dementia through the Use of Socially Assistive Robots	Music recognition (for games)	
2009	A. Tapus et al. [148]	The role of physical embodiment of a therapist robot for individuals with cognitive impairments	Music recognition (for games)	
2010	C. Kroos et al. [72]	The Articulated Head pays attention	Sound perception (loudness)	
2010	Y. Lin et al. [81]	Acoustical implicit communication in human-robot interaction	Paralanguage recognition	
2010	R. Nikolaidis and G. Weinberg [102]	Playing with the masters: A model for improvisatory musical interaction between robots and humans	Music synthesis (physical)	Instrumental music
2010	R. Read and T. Belpaeme [120]	Interpreting non-linguistic utterances by robots: studying the influence of physical appearance	Emotional robot sound Functional robot sound	Vocables; artificial sounds

Year	**Authors**	**Title**	**Function**	**Form & Techniques**
2010	H. A. Samani et al. [127]	Towards a formulation of love in human - robot interaction	Paralanguage recognition	
2010	R. K. Sarvadevabhatla et al. [130]	Extended duration human-robot interaction: Tools and analysis	Sound source localization	
2010	E. van der Heide [157]	Spatial Sounds (100dB at 100km/h) in the Context of Human Robot Personal Relationships	Transformative robot sound Functional robot sound	Electronic sounds
2011	H. Knight [69]	Eight Lessons Learned about Non-verbal Interactions through Robot Theater	Paralanguage recognition	
2011	A. Mertens et al. [94]	User focused design of human-robot interaction for people suffering from unusual ailments	Functional robot sound	Music; artificial sounds; natural sounds
2011	H. G. Okuno et al. [106]	Robot Audition: Missing Feature Theory Approach and Active Audition	Sound source localization Sound source separation	
2011	H. A. Samani et al. [126]	An affective interactive audio interface for Lovotics	Paralanguage recognition Emotional robot sound	Vocables
2011	M. Shiomi et al. [139]	Field Trial of a Networked Robot at a Train Station	Paralanguage recognition	
2011	J. P. Tissberger and G. Wersenyi [151]	Sonification Solutions for Body Movements in Rehabilitation of Locomotor Disorders	Functional robot sound	Music; sonification
2011	N. Yamakawa et al. [172]	Environmental Sound Recognition for Robot Audition Using Matching-Pursuit	Sound source recognition	
2012	T. Araki et al. [2]	Online Object Categorization Using Multimodal Information Autonomously Acquired by a Mobile Robot	Sound source recognition	
2012	K. S. Chun et al. [24]	Novel musical notation for Emotional robot sound expression of interactive robot	Sound creation (notation)	

Year	Authors	Title	Function	Form & Techniques
2012	G. Hoffman [54]	Dumb robots, smart phones: A case study of music listening companionship	Music recognition (for dance)	
2012	M. Janvier et al. [61]	Sound-event recognition with a companion humanoid	Sound source recognition	
2012	A. Lim et al. [79]	A Musical Robot that Synchronizes with a Coplayer Using Non-Verbal Cues	Music synthesis (physical) Music recognition	Instrumental music
2012	N. Masuyama et al. [87]	Computational Intelligence for Human Interactive Communication of Robot Partners	Sound source localization	
2012	J. L. Oliveira et al. [110]	An active audition framework for auditory-driven HRI: Application to interactive robot dancing	Music recognition (for dance)	
2012	J. S. Park et al. [111]	Music-aided affective interaction between human and service robot	Music recognition (for emotion) Paralanguage recognition	
2012	R. Read and T. Belpaeme [116]	How to use non-linguistic utterances to convey emotion in child-robot interaction	Emotional robot sound	
2013	G. Hoffman and K. Vanunu [56]	Effects of robotic companionship on music enjoyment and agent perception	Music recognition (for dance)	
2013	K. L. Koay et al. [70]	Exploring Robot Etiquette: Refining a HRI Home Companion Scenario Based on Feedback from Two Artists Who Lived with Robots in the UH Robot House	Functional robot sound	
2013	D. K. Limbu et al. [80]	Affective social interaction with CuDDler robot	Paralanguage recognition Emotional robot sound	
2013	S. Pourmehr et al. [115]	A robust integrated system for selecting and commanding multiple mobile robots	Emotional robot sound	

Year	Authors	Title	Function	Form & Techniques
2013	R. Read and T. Belpaeme [117]	People interpret robotic non-linguistic utterances Categorically	Emotional robot sound	Electronic sounds
2013	K. P. Tee et al. [149]	Audio-visual attention control of a pan-tilt telepresence robot	Sound source localization	
2013	A. Vasilijevic et al. [158]	Comparative assessment of human machine interfaces for ROV guidance with different levels of secondary visual workload	Functional robot sound	Auditory display; spatial sound
2013	J. von Zitzewitz et al. [160]	Quantifying the Human Likeness of a Humanoid Robot	Consequential robot sound	
2013	J. Y. Yang and D. Kwon [174]	Feedback-based reasoning process for behavior selection during long-term interaction	Emotional robot sound	
2014	S. Bökesoy [16]	A Recursive Mapping System For Motion And Sound In A Robot Between Human Interaction Design	Functional robot sound	Sonification; personalization
2014	K. Fischer et al. [31]	Initiating interactions in order to get help: Effects of social framing on people's responses to robots' requests for assistance	Functional robot sound	Electronic sounds
2014	K. Fischer et al. [30]	To Beep or Not to Beep Is Not the Whole Question	Functional robot sound	Vocables; music
2014	S. E. Fotinea et al. [33]	The annotation scheme of the MOBOT dataset	Sound source localization	
2014	M. Janvier et al. [62]	Sound representation and classification benchmark for domestic robots	Sound source recognition	
2014	M. Joosse et al. [67]	Sound over matter: the effects of functional noise, robot size and approach velocity in human-robot encounters	Transformative robot sound	

Year	Authors	Title	Function	Form & Techniques
2014	E. Martinson and V. Yalla [85]	Guiding computational perception through a shared auditory space	Sound source localization	
2014	L. McCallum and P. W. McOwan [88]	Shut up and play: A musical approach to engagement and social presence in Human Robot Interaction	Music synthesis (physical)	Instrumental music
2014	R. Read and T. Belpaeme [118]	Non-linguistic utterances should be used alongside language, rather than on their own or as a replacement	Emotional robot sound	Vocables
2014	R. Read and T. Belpaeme [119]	Situational context directs how people affectively interpret robotic non-linguistic utterances	Emotional robot sound	Vocables
2014	M. Schwenk and K. O. Arras [133]	R2-D2 Reloaded: A flexible sound synthesis system for sonic human-robot interaction design	Emotional robot sound Functional robot sound	Sonification
2014	F. Speth and M. Wahl [142]	Specifying Rhythmic Auditory Stimulation for Robot-assisted Hand Function Training in Stroke Therapy	Functional robot sound	Music
2014	R. Stęgierski and K. Kuczyński [144]	The Perception of Humanoid Robot by Human	Sound source localization	
2015	L. Boccanfuso et al. [8]	Autonomously detecting interaction with an affective robot to explore connection to developmental ability	Emotional robot sound	Vocables; music
2015	G. Ince et al. [57]	Towards a robust drum stroke recognition system for human robot interaction	Music synthesis (physical) Music recognition	Instrumental music
2015	L. McCallum and P. W. McOwan [89]	Face the Music and Glance: How Nonverbal Behaviour Aids Human Robot Relationships Based in Music	Music synthesis (physical)	Instrumental music
2015	H. Peng et al. [113]	Robotic Dance in Social Robotics—A Taxonomy	Music recognition (for dance)	

Year	Authors	Title	Function	Form & Techniques
2015	E. Sandry [128]	Re-evaluating the Form and Communication of Social Robots: The Benefits of Collaborating with Machinelike Robots	Emotional robot sound	Vocables
2016	A. F. Azmin et al. [3]	HRI observation with My Keepon robot using Kansei Engineering approach	Music recognition (for dance) Emotional robot sound	
2016	E. Cha et al. [21]	Nonverbal signaling for non-humanoid robots during human-robot collaboration	Functional robot sound	
2016	E. Cha and M. Matarić [22]	Using nonverbal signals to request help during human-robot collaboration	Functional robot sound	Electronic sounds
2016	E. Florentine et al. [32]	Pedestrian Notification Methods in Autonomous Vehicles for Multi-Class Mobility-on-Demand Service	Functional robot sound	Music
2016	T. Giannakopoulos and G. Siantikos [38]	A ROS framework for audio-based activity recognition	Sound source recognition	
2016	H. Hastie et al. [50]	Sound emblems for affective multimodal output of a robotic tutor: a perception study	Emotional robot sound	Vocables
2016	G. Hoffman et al. [55]	Robotic experience companionship in music listening and video watching	Music recognition (for dance)	
2016	H. Kudo et al. [74]	Behavior Model for Hearing-Dog Robot	Sound source recognition	
2016	S. Lakhmani et al. [75]	A Proposed Approach for Determining the Influence of Multimodal Robot-of-Human Transparency Information on Human-Agent Teams	Functional robot sound	Music
2016	M. C. Shrestha et al. [140]	Exploring the use of light and display indicators for communicating directional intent	Functional robot sound	
2016	A. Taheri et al. [145]	Social Robots and Teaching Music to Autistic Children: Myth or Reality?	Music synthesis (physical) Music recognition	Instrumental music

Year	Authors	Title	Function	Form & Techniques
2016	M. Tahon and L. Devillers [146]	Towards a Small Set of Robust Acoustic Features for Emotion Sound source recognition: Challenges	Paralanguage recognition	
2016	G. Xia et al. [170]	Expressive Humanoid Robot For Automatic Accompaniment	Music synthesis (physical) Music recognition (for games)	Instrumental music
2016	S. Yilmazyildiz et al. [175]	Review of Semantic-Free Utterances in Social Human–Robot Interaction	Emotional robot sound Functional robot sound	Music; vocables
2016	C. Zaga et al. [176]	Help-giving robot behaviors in child-robot games: Exploring Semantic Free Utterances	Functional robot sound	Vocables
2016	R. Zhang et al. [181]	Musical Robots For Children With ASD Using A Client-Server Architecture	Emotional robot sound	Music; sonification
2017	R. Agrigoroaie and A. Tapus [1]	Influence of Robot's Interaction Style on Performance in a Stroop Task	Sound creation	Digital sounds
2017	J. Bellona et al. [6]	Empirically Informed Sound Synthesis Application for Enhancing the Perception of Expressive Robotic Movement	Transformative robot sound	Sonification
2017	L. Dahl et al. [26]	Data-Driven Design of Sound for Enhancing the Perception of Expressive Robotic Movement	Transformative robot sound	Sonification
2017	J. Fernandez De Gorostiza luengo et al. [29]	Sound Synthesis for Communicating Nonverbal Expressive Cues	Emotional robot sound	Music Electronic sounds
2017	E. Jeong et al. [64]	Exploring the taxonomic and associative link between emotion and function for robot sound design	Emotional robot sound Functional robot sound	Vocables; Electronic sounds
2017	D. Moore et al. [98]	Making Noise Intentional: A Study of Servo Sound Perception	Consequential robot sound	Mechanical sounds
2017	E. Sandry [129]	Creative Collaborations with Machines	Music synthesis (virtual) Music recognition (for dance)	

Year	Authors	Title	Function	Form & Techniques
2017	M. Shahab et al. [134]	Social Virtual Reality Robot (V2R): A Novel Concept for Education and Rehabilitation of Children with Autism	Music synthesis (virtual) Music recognition (for games)	Music
2017	H. Tennent et al. [150]	Good vibrations: How consequential sounds affect perception of robotic arms	Consequential robot sound	Mechanical sounds
2018	G. Bolano et al. [9]	Transparent Robot Behavior by Adding Intuitive Visual and Acoustic Feedback to Motion Replanning	Functional robot sound	
2018	E. Cha et al. [20]	Effects of Robot Sound on Auditory Sound source localization in Human-Robot Collaboration	Transformative robot sound	Electronic sounds
2018	E. Frid et al. [36]	Perception Of Mechanical Sounds Inherent To Expressive Gestures Of A Nao Robot - Implications For Movement Sonification Of Humanoids	Consequential robot sound Emotional robot sound	
2018	L. Grama and C. Rusu [45]	Adding audio capabilities to TIAGo service robot	Sound source localization Sound source recognition	
2018	W. He et al. [51]	Deep Neural Networks for Multiple Speaker Detection and Sound source localization	Sound source localization	
2018	D. Löffler et al. [82]	Multimodal Expression of Artificial Emotion in Social Robots Using Color, Motion and Sound	Emotional robot sound	Electronic sounds
2018	L. McCallum and P. W. McOwan [90]	Extending Human–Robot Relationships Based in Music With Virtual Presence	Music synthesis (physical)	Instrumental music
2018	A. Nijholt [101]	Robotic Stand-Up Comedy: State-of-the-Art	Sound perception (loudness)	
2018	K. Shibuya and H. Ishimoto [138]	Design Principles of Loudness to Express Bright and Dark Timbres for Violin-playing Robot	Music synthesis (physical)	Instrumental music
2018	G. Trovato et al. [153]	The Sound or Silence: Investigating the Influence of Robot Noise on Proxemics	Transformative robot sound	Music

Year	Authors	Title	Function	Form & Techniques
2018	K. Weber et al. [161]	How to Shape the Humor of a Robot - Social Behavior Adaptation Based on Reinforcement Learning	Sound perception (loudness) Emotional robot sound	Personalization
2018	K. Weber et al. [162]	Real-Time Adaptation of a Robotic Joke Teller Based on Human Social Signals	Sound perception (loudness) Emotional robot sound	
2019	M. R. Frederiksen and K. Stoey [34]	Augmenting the audio-based expression modality of a non-affective robot	Emotional robot sound	Vocables
2019	P. Jin et al. [65]	A-EXP4: Online Social Policy Learning for Adaptive Robot-Pedestrian Interaction	Functional robot sound	
2019	L. Martínez-Villaseñor and H. Ponce [86]	A concise review on sensor signal acquisition and transformation applied to human activity recognition and human–robot interaction	Sound source recognition	
2019	H. Ritschel et al. [121]	Personalized Synthesis of Intentional and Emotional Non-Verbal Sounds for Social Robots	Functional robot sound	Music; personalization
2019	S. Rossi et al. [124]	Evaluating the Emotional Valence of Affective Sounds for Child-Robot Interaction	Emotional robot sound	Vocables
2019	R. Savery et al. [132]	Establishing Human-Robot Trust through Music-Driven Robotic Emotion Prosody and Gesture	Emotional robot sound	Vocables
2019	A. Ueno et al. [156]	Impression Change on Nonverbal Non-Humanoid Robot by Interaction with Humanoid Robot	Functional robot sound	Electronic sounds
2020	L. Boos and L. Moshkina [11]	Conveying Robot State and Intent Nonverbally in Military-Relevant Situations: An Exploratory Survey	Consequential robot sound Functional robot sound	Vocables Electronic sounds
2020	S. Chakraborty and J. Timoney [23]	Robot Human Synchronization for Musical Ensemble: Progress and Challenges	Music synthesis (physical) Music recognition	Instrumental music

Year	Authors	Title	Function	Form & Techniques
2020	W. K. N. Hansika et al. [48]	AuDimo: A Musical Companion Robot to Switching Audio Tracks by Recognizing the Users Engagement	Music recognition (for dance)	
2020	T. Izui and G. Venture [59]	Correlation Analysis for Predictive Models of Robot User's Impression: A Study on Visual Medium and Mechanical Noise	Consequential robot sound	
2020	S. Jaiswal et al. [60]	Image based Emotional State Prediction from Multiparty Audio Conversation	Paralanguage recognition	
2020	A. B. Latupeirissa et al. [77]	Exploring emotion perception in sonic HRI	Emotional robot sound	Electronic sounds
2020	J. Okimoto and N. Niitsuma [103]	Effects of Auditory Cues on Human-Robot Collaboration	Functional robot sound	Electronic sounds
2020	H. R. M. Pelikan et al. [112]	"Are You Sad, Cozmo?": How Humans Make Sense of a Home Robot's Emotion Displays	Emotional robot sound	Vocables
2020	J. Vilk and N. T. Fitter [159]	Comedians in Cafes Getting Data: Evaluating Timing and Adaptivity in Real-World Robot Comedy Performance	Paralanguage recognition	
2020	H. Wolfe et al. [169]	Singing Robots: How Embodiment Affects Emotional Responses to Non-Linguistic Utterances	Emotional robot sound	Computer music
2021	J. A. Barnes et al. [4]	Child-Robot Interaction in a Musical Dance Game: An Exploratory Comparison Study between Typically Developing Children and Children with Autism	Music recognition (for dance) Music recognition (for games)	
2021	F. Ciardo et al. [25]	Effects of erring behavior in a human-robot joint musical task on adopting Intentional Stance toward the iCub robot	Music synthesis (physical) Music recognition (for games)	Instrumental music

Year	Authors	Title	Function	Form & Techniques
2021	J. Fan et al. [28]	Field Testing of Ro-Tri, a Robot-Mediated Triadic Interaction for Older Adults	Sound source localization	
2021	G. Ince et al. [58]	An audiovisual interface-based drumming system for multimodal human–robot interaction	Music synthesis (physical) Music recognition (for games)	Instrumental music
2021	M. Krzyżaniak [73]	Musical robot swarms, timing, and equilibria	Music synthesis (physical) Music recognition (for games)	Instrumental music
2021	J. S. Lee et al. [78]	Non-Verbal Auditory Aspects of Human-Service Robot Interaction	Transformative robot sound Emotional robot sound Functional robot sound	
2021	L. Muscar et al. [99]	Sound Classification by the TIAGo Service Robot for Healthcare Applications	Sound source recognition	
2021	T. R. P. Pessanha et al. [114]	Virtual Robotic Musicianship: Challenges and Opportunities	Music synthesis	Instrumental music
2021	R. Savery et al. [131]	Emotion Musical Prosody for Robotic Groups and Entitativity	Emotional robot sound	Vocables
2021	S. C. Steinhaeusser et al. [143]	Comparing a Robotic Storyteller versus Audio Book with Integration of Sound Effects and Background Music	Emotional robot sound	Music
2021	B. J. Zhang et al. [180]	Bringing WALL-E out of the Silver Screen: Understanding How Transformative Robot Sound Affects Human Perception	Transformative robot sound Emotional robot sound	Vocables; mechanical sound; electrical sound
2021	B. J. Zhang et al. [178]	Exploring Consequential Robot Sound: Should We Make Robots Quiet and Kawaii-et?	Consequential robot sound	

Year	Authors	Title	Function	Form & Techniques
2022	E. Frid and R. Bresin [35]	Perceptual Evaluation of Blended Sonification of Mechanical Robot Sounds Produced by Emotionally Expressive Gestures: Augmenting Consequential Sounds to Improve Non-verbal Robot Communication	Consequential robot sound Transformative robot sound	Computer music
2022	M. A. Maheux et al. [83]	T-Top, a SAR Experimental Platform	Sound source localization Sound source recognition Sound source separation	
2022	U. Maniscalco et al. [84]	Bidirectional Multi-modal Signs of Checking Human-Robot Engagement and Interaction	Sound source localization Sound perception (loudness) Functional robot sound	Electronic sounds
2022	M. Shahab et al. [135]	Utilizing social virtual reality robot (V2R) for music education to children with high-functioning autism	Music synthesis (virtual) Music recognition (for games)	Instrumental music

Table 1. Articles included in the review, arranged by year of publication and alphabetical order of the first author surname. The “Function” column labels each article with a relevant topic from the taxonomy of function described in Section 3.1, while the “Form & Technique” column indicates the type of sound creation in relevant papers that sufficiently describe the sound. Table 4, which organizes these works by “Function,” is located in the appendix.

3.1 Terms and Taxonomies

The words used to categorize nonverbal sound in human-robot interaction have varied greatly between works. Some authors have borrowed from adjacent fields, such as music, product sound design, auditory display, and computational linguistics; other authors have created terms that they believe best suit their topic of study. The quest to taxonomize nonverbal sound poses extra difficulty due to the wide range of forms that sound may take. However, a lack of common terms also creates problems for research, as descriptions of design and implementation methods for nonverbal sound do not effectively enable other researchers to replicate prior work. Alternatives for sharing research, such as providing software or sound files, remain rare; many papers on sound do not provide such files, provide files that have since become unavailable, or only provide such files in a form that includes environmental background noise (e.g., in video-based stimuli). While these issues may be alleviated by increasing trends of including open-source tools and multimedia attachments with academic works, researchers will benefit from an imminent common and accessible set of terms for the field.

We examined prior efforts to taxonomize nonverbal sound in human-robot interaction with a particular emphasis on scientific communication, as this field lies at the intersection of several disciplines such as engineering, social science, and musicology. Thus, terms have a particular risk of becoming *jargon*, as terms with roots from one discipline may act as jargon to another. Based on these prior efforts and the authors' experiences, we developed new taxonomies for form and function, with associated recommendations on using these new terms for future research.

3.1.1 Previous Taxonomies. We searched the reviewed articles for explicit taxonomies of nonverbal sound, the results of which are presented in Table 2. Taxonomies ranged in purpose from categorizing nonverbal sound in the context of a study to categorizing the topics of articles within a literature review. We considered each taxonomy within the context of broader taxonomies for sound, such as from psychoacoustics [10, 37], product sound design [76], and auditory display [52]. The taxonomies generally fell into two categories: *form*, where nonverbal sounds were categorized based on how they sound, and *function*, where nonverbal sounds were categorized based on their purpose.

Form-based taxonomies offer visual information on the auditory nature of a sound; they describe what the sound sounds like. These taxonomies can accomplish this task in several ways; one such way is by through associations. For instance, Mertens et al. uses sounds from Microsoft Windows [94], which provides a suite of *earcons* ("a brief,

Authors	**Categorization**
Mertens et al. [94]	Everyday, nature, Microsoft Windows, jingles
Janvier et al. [61]	Prosodic, moving, cooking, alarms
Janvier et al. [62]	Kitchen, office, nonverbal, speech
Yilmazyildiz et al. [175]	Semantic-free utterances (gibberish speech, paralinguistic utterances, musical utterances, non-linguistic utterances)
Jeong et al.[64]	Robot sound design (functional (platform, monitoring, alerting, feedback), emotional (positive, neutral, negative))
Lee et al. [78]	Audible communication (auditory icons/earcons, ambient background sound, anthropomorphic intent notifiers)

Table 2. Articles that propose a taxonomy for nonverbal robot sound, with a brief description of the structure of the taxonomy.

distinctive sound that represents a specific item or event" [163]) that may be easily recalled by Microsoft Windows users. Other associations, such Janvier et al.'s cooking and kitchen, may also lead to common understandings [61, 62]. Sound taxonomies from psychoacoustics do already offer a more comprehensive set of associations [10]. However, a key weakness of categorizing sounds through association becomes broader and more esoteric categories. Descriptors such as *nature* and *office* may have variable meaning depending on the reader's geographical location and place of work. In each of these articles, the authors further describe each sound to help account for this concern, though descriptions of form can still suffer from lack of clarity. Mertens et al. also categorized sounds as *jingles*, "memorable short song[s], or in some cases a snippet of a popular song" [164], that include too many sounds for complete descriptions. Overall, taxonomies of form using associations alone do not provide a clear and consistent depiction of sounds.

Yilmazyildiz et al.'s previous review proposed a taxonomy of form based on definitions rather than association. This taxonomy provides an alternative, more specific term for nonverbal sound: semantic-free utterances. Semantic-free utterances is divided into four forms: *gibberish speech* ("vocalizations of meaningless strings of speech sounds"), *paralinguistic utterances* ("stand-alone vocal events"), *musical utterances* (music and music theory-based sounds), and *non-linguistic utterances* (other nonspeech-like sounds) [175]. More generally, *paralanguage* is "the non-verbal elements of speech...such as pitch, volume, and intonation" [165]. Of these forms, non-linguistic utterances has seen the most use, with roots from the works of two of Yilmazyildiz's co-authors, Read and Belpaeme. Read and Belpaeme coined non-linguistic utterances, though in older work it also included terms might be categorized as gibberish speech or paralinguistic utterances, such as human nonverbal utterances [120] and sounds produced by the characters Chewbacca and WALL-E [116]. A more specific definition for non-linguistic utterances was provided in [118]: "robotic sounds made by synthetic social agents, rather than utterances that are designed to resemble natural speech, such as artificial languages or gibberish speech."

As the most established taxonomy, semantic-free utterances and its components have been referenced in several of the reviewed articles [36, 36, 64, 124, 169, 176]. However, the terms have already experienced deviation from their intended use. Small deviations include changes to the terms while maintaining the intended designations, such as changing non-linguistic utterances to "non-linguistic functional sound" [64] or "non-linguistic auditory cues" [169] and changing paralinguistic utterances to "para-linguistic vocalizations" [124] or simply "paralanguage" [169]. Larger deviations stem from uncertainty over how speech-like or musical non-linguistic utterances can be, which may cause miscategorization of sounds [169]. These deviations point toward a root issue: the use of jargon. This taxonomy primarily uses terms from linguistics (semantic-free, paralinguistic, non-linguistic), which provides great specificity in at the cost of inaccessibility to readers unfamiliar with linguistics. Furthermore, these multi-part terms often become abbreviations (semantic-free utterance as SFU, gibberish speech as GS, etc.). Abbreviations are also a form of jargon that can be particularly detrimental to readers. With the exception of the most widely-known abbreviations (e.g., SCUBA, LASER), readers must parse and re-reference abbreviations throughout the text to reconstruct their meaning; readers meeting unfamiliar abbreviations may also feel alienated and less interested [47]. These concerns spurred us to pursue updated terms.

Function-based taxonomies, on the other hand, offer information on the intended purpose of sounds, which may help researchers connect and collaborate over similar goals. Lee et al. propose a taxonomy that blends both form and function, separating nonverbal robot sound into auditory icons or earcons ("cues, notifications, informational alerts, feedback"), ambient background sound ("to indicate that a robot is nearby, or to establish mood and situation"), and anthropomorphic intent notifiers ("specifically to relate to humans in the vicinity") [78]. While each category contains information on form, the categorization distinguishes itself with different functions. However, the taxonomy is not as comprehensive when compared to Yilmazyildiz et al., as the the combination of form and function excludes certain combinations. For instance, some forms of sonification (the process of "map[ping] data to sound" [166]) may lead to ambient background sound that provides informational alerts or feedback, creating a cross-category sound. Directly combining form and function may lead to these types of discrepancies.

Jeong et al. propose a purely function-based taxonomy with a simple separation: emotion and functional sound [64]. This article notes an important concern for taxonomies of function: sounds designed for a particular purpose may be interpreted by the listener differently or in more dimensions than intended. In particular, functional sounds may have emotional content. However, for individuals researching nonverbal sound, the intended function is more important to indicate, as the resulting effect can be framed as a measure of the sound's effectiveness in its intended function. Another concern for Jeong et al.'s taxonomy is that these nonverbal sounds do not encompass the entirety of nonverbal robot sound, as robots also produce consequential sound, or sound generated by the operating of the robot itself [76]. These sounds may not serve any particular emotional or functional role, but still may be designed and affect the human-robot interactions. Furthermore, the addition of transformative sound, intentionally produced sound intended to alter a robot's original sound profile, may combine with consequential sound as an alternative to produce a new overall sound profile without changes to the physical design of the robot [180].

Overall, taxonomies of form and function serve useful but separate purposes. The existing taxonomies leave some concerns, particularly as few works succeed in offering both types of categorization. Thus, we developed new taxonomies for the field that aim to alleviate these concerns.

3.1.2 Proposed Taxonomy. We propose new taxonomies of form and function, designed to be used in conjunction with one another to provide clear descriptions of nonverbal sound for future articles. Firstly, a new taxonomy of function for sound, shown in Figure 2, identifies the current major research focuses found through the systematic review. This taxonomy provides overarching structure by dividing the role of sound relative to a robot into *sound perception* and *sound creation*, which are further divided into *implicit* and *explicit*.

Implicit and explicit perception carry different meanings than implicit and explicit creation. Explicit sound perception concerns the properties of the sound itself, such as sound pressure level, frequencies, duration, or location, while implicit sound perception uses the sound to infer characteristics of something else, such as the object or interaction producing the sound. On the other hand, implicit and explicit in sound creation follows the convention of implicit and explicit communication in human-robot interaction; explicit sound creation deliberately conveys information with a clear associated intent for the listener to receive said information, while implicit sound creation does not necessarily convey information, but the listener may infer information from the sound anyways [39]. Rather than binary classification, the implicit and explicit categorization should be viewed more as a continuous scale, as shown in Figure 2, and functions of sound also do not occupy a single point on the scale. To illustrate this, we include speech within the taxonomy alongside paralanguage, as all speech contains paralanguage. While speech recognition (still beyond the scope of the review, but included as an easy-to-understand reference point) is a clear example of implicit sound perception, as it focuses on the linguistic representation of sounds and their meaning, paralanguage recognition is a clear example of explicit sound perception, as it focuses on auditory features such as loudness and pitch, and so the two concepts combined fall at the center of the scale as the concepts both apply to perception of speech. Similarly, transformative robot sound may only implicitly communicate, such as by simply amplifying consequential robot sounds to increase the noticeability of the robot, or be designed to explicitly communicate, as is the intent of blended sonification ("sonifications that blend into the users' environment without confronting users with any explicitly perceived technology" [155]). Thus, these implicit

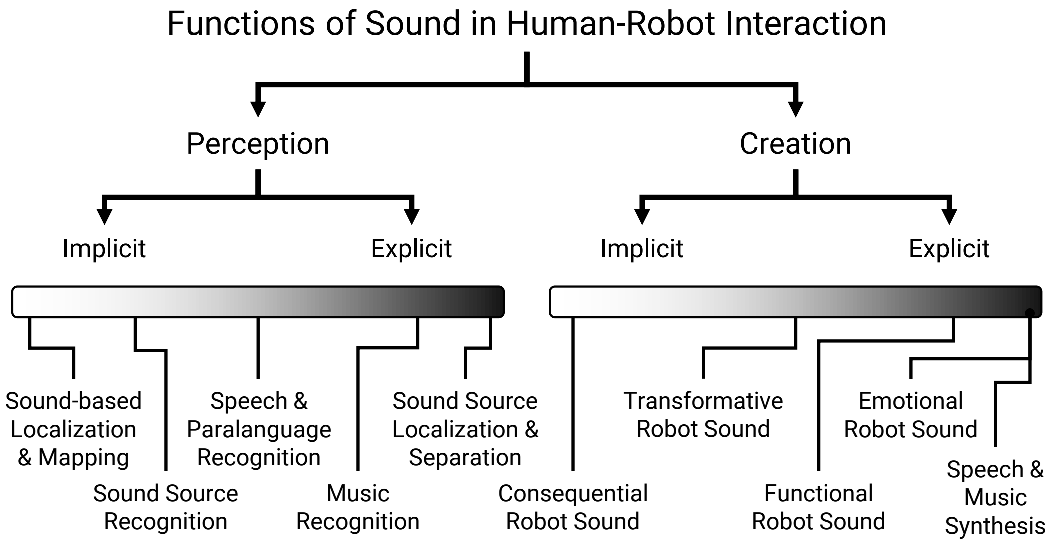

Fig. 2. The taxonomy of function for sound in human-robot interaction, with current major functions placed within the taxonomy. This taxonomy is robot-centric, or from the perspective of the robot; that is, "perception" refers to the robot's perception of sound and "creation" refers to sounds that the robot creates. The location of each function on the implicit-explicit scales acts as a general guideline rather than a specific categorization.

to explicit scales can accommodate novel functions of sound in human-robot interaction, placing them in a holistic picture of the field.

From the reviewed articles, we identified the following functions and general definitions:

- **Sound source recognition**: identifying the objects and interactions that produce sounds in a robot's environment.
- **Paralanguage recognition**: extracting information from speech based on paralanguage or identifying nonverbal vocal sounds.
- **Music recognition**: extracting musical information from sound, such as instruments, notes, and tempo.
- **Sound source localization**: identifying the location of origin of a sound in a robot's environment.
- **Sound source separation**: identifying that sounds have different origins or splitting the audio signal of sounds with different origins.
- **Consequential robot sound**: sound made by the operation of the robot.
- **Transformative robot sound**: sound made to mix with or act as consequential robot sound with the intent to change the sound profile of the robot.
- **Functional robot sound**: sound made to explicitly convey non-emotional information from the robot.
- **Emotional robot sound**: sound made to explicitly convey emotions from the robot.
- **Music synthesis**: the creation of music either through electronics or through physical instruments.

While most of the sound perception categories are firmly established, the categorization of sound creation arises from a combination of Jeong et al.'s taxonomy and product sound design. As previously mentioned, Jeong et al. proposed a taxonomy of emotional and functional sound [64]. Jeong et al.'s taxonomy accommodates sound for explicit

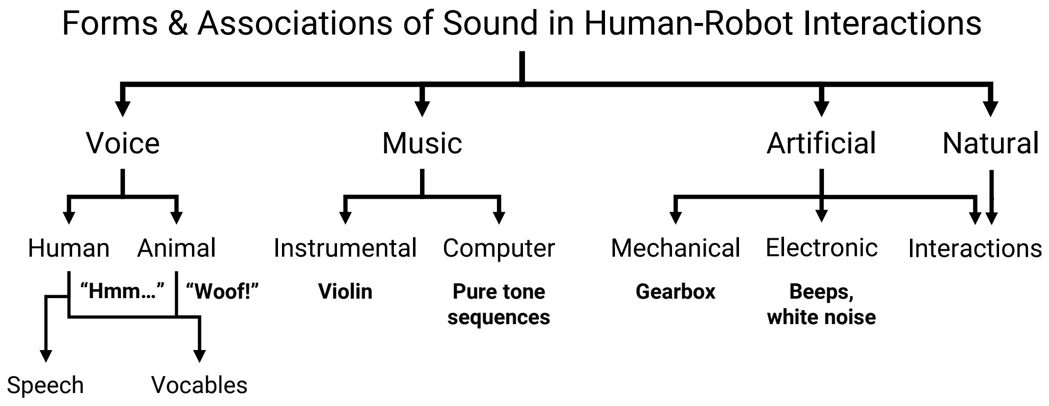

Fig. 3. The taxonomy of form for sound in human-robot interaction, with several examples of sounds and sound origins. Some descriptors, such as "vocables," can be used separately, while others must be combined with "sound" as in "mechanical sound" or with each other as in "computer music."

communication, but omits implicit created sounds such as consequential sound from product sound design [76], which has been examined in several of the reviewed articles. Furthermore, the addition of consequential sound provides an opportunity to differentiate emotional and functional sound further, as sounds created to complement, mask, or otherwise *transform* consequential sound can be separated into transformative robot sound, also a topic found within the review.

Figure 2 further includes sound-based localization and mapping, the use of sound to determine the position of the robot or objects in the environment. Sound-based localization and mapping often uses ultrasound (sound above the range of human hearing) and is thus out of the scope of sound in human-robot interaction. We include sound-based localization and mapping in the overall taxonomy for reference. Table 4 in the appendix organizes the reviewed articles by function, rather than chronological order.

The taxonomy of form complements the taxonomy of function for articles on sound creation by providing more specific descriptions of sounds. As with the taxonomy of function, we examined terms currently in use from the perspective of scientific communication. The taxonomy of form, presented in Figure 3, combines and rearranges sound taxonomies [10, 37] to better suit robot sound and adds *vocable* ("(linguistics) a word or utterance especially with reference to its form rather than its meaning; (music) a syllable or sound without specific meaning, used together with or in place of actual words in a song" [167]) as an alternative for gibberish speech and paralinguistic utterances from Yilmazyildiz et al.'s taxonomy [175].

The introduction of *vocable* stems from the need for a term to describe vocalizations that do not form words or speech. Previous candidates have included Yilmazyildiz et al.'s gibberish speech, paralinguistic utterances, and non-linguistic utterances [175]. However, as previously noted, these terms best suit individuals with a background in linguistics and may act as jargon for others, particularly when abbreviated. In comparison, vocable shares etymological roots with common words such as vocal and vocabulary (*voco* from Latin, meaning "I call" [167]) and is short enough to not be abbreviated. Lastly, vocable is an established term with greater popularity in both English books and global search trends according to Google Books Ngram Viewer and Google Search Trends [40, 42]. The term "vocables" correlates most closely to paralinguistic utterances and may also be used to describe animal vocalizations ("animal vocables").

Gibberish speech may also be described as vocables, though gibberish speech is composed of a series of vocables; thus, the "gibberish speech" term can provide additional detail if desired.

Another key change in the taxonomy concerns the use of *music* and the exclusion of *musical*. This distinction is important, as many non-music sounds still draw inspiration from music theory. As an example, Fischer et al. created sounds based on the spoken phrase "excuse me, please" with different intonation. Fischer et al. converted the intonation into beeps (electronic sounds) and presented the beeps in music notation and noted their musical relationship in terms of semitones [30]. Thus, these sounds could be thought of as musical, but were intended to emulate speech, not music. In the presented taxonomy, Fischer et al.'s sounds would be categorized as electronic sounds, which most closely match the final product. We recommend describing robot sound as music primarily for instrumental music and extended sequences of music theory-based electronic sounds.

Noise ("sound, especially one that lacks an agreeable quality or is noticeably unpleasant or loud; any sound that is undesired or interferes with one's hearing of something" [91]) may also cause confusion when used to describe useful sound (e.g., "functional noise" [67]). We recommend that sounds be described as noise only when undesirable or in reference to specific forms of noise, such as white, pink, or Brownian noise, terms established in signal processing that also describe audio signals [168]. Another use of the term appears in *ego-noise*, the undesired consequential sound when considering sound perception [110]. While self-noise (the audio signal measured by a microphone that is not caused by other sound sources [100]) is more popular that ego-noise in both English books and global search trends according to Google Books Ngram Viewer and Google Search Trends [41, 43], ego-noise seems established in the literature for sound perception by robots.

Natural sounds refer to sounds that, regardless of production method, evoke nature. For instance, water drops or pouring, footsteps on grass, and the crackling of leaves or fire all fall under natural sounds. Natural sounds do not appear frequently in the literature, but may feature more widely as robots increasingly enter the outdoors. At the time, it may be valuable to introduce further distinctions within this category of sound.

3.1.3 Usage Recommendations. The taxonomy of function can help researchers find articles with similar goals. Articles on nonverbal sound in human-robot interaction should mention the associated function(s) of sound in unabbreviated form as early and often as possible, preferably within the title or abstract of the work. Within the full text of the paper, limited shortening is reasonable. For instance, an article on sound source localization that does not include other localization topics, such as simultaneous localization and mapping, may reasonably shorten the full function name to "sound localization" or simply "localization." In the same vein, transformative robot sound may be condensed to "transformative sound." We strongly recommend against the use of abbreviations into acronyms and initialisms, which often require readers to repeatedly find the first mention of the relevant term and alienate readers new to the field [47]. When acronyms or initialisms are required for space, such as within a figure, figure captions should link the full term and the abbreviation together. Until the research community of robot sound establishes a common vocabulary, these steps are essential for unifying the field.

The taxonomy of form should be used for concise description of sounds. Similarly to the taxonomy of function, articles on robot sound creation should mention the associated form(s) of sound as early as possible, preferably in the title or abstract of the work. Within the full text of the paper, the taxonomy of form should be used for quick references to the sounds, such as to differentiate between two sound designs (e.g., to contrast sound designs using vocables or electronic sound). However, describing sounds with this taxonomy does not replace a more detailed explanation of the sound design methods and form. We strongly recommend that authors make created robot sounds available to readers in raw audio form, in addition to in the context of study stimuli.

3.2 Study Methods for Nonverbal Sound Creation

As investigations into nonverbal sound creation have progressed, study methods have correspondingly evolved. In this work, we aimed to provide an updated view of study methods for nonverbal sound creation from Yilmazyildiz et al.'s review. Instead of reporting the same metrics, we focused on the study designs. Table 3 details the results of our search.

Year	**Title**	**Robot**	**Measures**	**Statistical Testing**	**Sound Comparison**	**Multimodal Comparison**	**Study Design**
2002	Auditory display of directions and states for mobile systems [66]	None	Objective & subjective	No	Sound case studies; Between sounds	No	In-person; computer-based; within-subjects
2003	Interactive visualization and sonification for monitoring complex processes [53]	Robot arm	Objective & subjective	No	Between sounds	No	In-person; computer-based; within-subjects
2007	Action Sloping as a Way for Users to Notice a Robot's Function [71]	AIBO	Objective	Yes	Presence of sound	Yes	In-person; experimental; between-subjects
2007	Analysis by Synthesis of an Information Presentation Method of Embodied Agent Based on the Time Lag Effects of Utterance to Communicative Actions [173]	None	Subjective	No	Between sounds	No	In-person; experimental; within-subjects
2010	Interpreting Non-Linguistic Utterances by Robots: Studying the Influence of Physical Appearance [120]	NAO, AIBO	Subjective	Yes	Between sounds	No	Online survey; image-based AND sound-based; within-subjects
2011	An Affective Interactive Audio Interface for Lovotics [126]	Lovotics	Subjective	Yes	Between sounds	No	In-person; experimental; within-subjects
2011	User focused design of human-robot interaction for people suffering from unusual ailments [94]	None	Objective & subjective	Yes	Between sounds	No	In-person; experimental; within-subjects

Year	Title	Robot	Measures	Statistical Testing	Sound Comparison	Multimodal Comparison	Study Design
2012	How to use non-linguistic utterances to convey emotion in child-robot interaction [116]	NAO	Subjective	No	Between sounds	No	In-person; experimental; within-subjects
2013	Comparative assessment of human machine interfaces for ROV guidance with different levels of secondary visual workload [158]	None	Objective	No	Presence of sound	Yes	In-person; experimental; within-subjects
2013	People interpret robotic non-linguistic utterances Categorically [117]	NAO	Subjective	Yes	Between sounds	No	In-person; experimental; within-subjects
2013	Quantifying the Human Likeness of a Humanoid Robot [160]	Bioloid	Subjective	No	Presence of sound	No	Online survey; video-based; within-subjects
2014	Initiating interactions in order to get help: Effects of social framing on people's responses to robots' requests for assistance [31]	PR2	Objective & subjective	Yes	Between verbal and nonverbal	No	In-person; experimental; between-subjects
2014	Non-Linguistic Utterances Should be Used Alongside Language, Rather than on their Own or as a Replacement [118]	NAO	Subjective	Yes	Between verbal and nonverbal	No	Online survey; video-based; within-subjects
2014	Situational Context Directs How People Affectively Interpret Robotic Non-Linguistic Utterances [119]	NAO	Subjective	Yes	Between sounds, including none	No	Online survey; video-based; within-subjects

Year	**Title**	**Robot**	**Measures**	**Statistical Testing**	**Sound Comparison**	**Multimodal Comparison**	**Study Design**
2014	Sound over Matter: The Effects of Functional Noise, Robot Size and Approach Velocity in Human-Robot Encounters [67]	Giraff	Subjective	Yes	Between sounds	No	In-person; experimental; between-subjects
2014	Specifying Rhythmic Auditory Stimulation for Robot-assisted Hand Function Training in Stroke Therapy [142]	None	Subjective	Yes	Between sounds, including none	No	In-person; experimental; within-subjects
2014	To Beep or Not to Beep Is Not the Whole Question [30]	Care-O-bot	Subjective	Yes	Between sounds, including none	No	In-person; experimental; between-subjects
2016	Exploring the use of light and display indicators for communicating directional intent [140]	Custom mobile robot	Subjective	Yes	Presence of sound	Yes	In-person; experimental; within-subjects
2016	Help-giving robot behaviors in child-robot games: Exploring Semantic Free Utterances [176]	Festo Robotino	Objective	Yes	Between sounds, including none	No	In-person; experimental; between-subjects
2016	Sound Emblems for Affective Multimodal Output of a Robotic Tutor: A Perception Study [50]	NAO	Subjective	Yes	Between sounds, including none	No	Online survey; sound-based; mixed methods
2016	Using nonverbal signals to request help during human-robot collaboration [22]	Ava	Objective & subjective	Yes	Between sounds, including none	No	In-person; experimental; within-subjects
2017	Data-Driven Design of Sound for Enhancing the Perception of Expressive Robotic Movement [26]	None	Subjective	Yes	Between sounds	No	Authors; computer-based; within-subjects

Year	Title	Robot	Measures	Statistical Testing	Sound Comparison	Multimodal Comparison	Study Design
2017	Exploring the taxonomic and associative link between emotion and function for robot sound design [64]	None	Subjective	Yes	Between sounds	No	In-person; computer-based; within-subjects
2017	Good vibrations: How consequential sounds affect perception of robotic arms [150]	youBot, OWI	Subjective	Yes	Between sounds, including none	No	Online survey; video-based; within-subjects
2017	Making Noise Intentional: A Study of Servo Sound Perception [98]	None	Subjective	Yes	Between sounds	No	Online survey; sound-based; mixed methods
2017	Sound Synthesis for Communicating Nonverbal Expressive Cues [29]	None	Subjective	No	Between sounds	No	Online survey; sound-based; within-subjects
2018	Effects of Robot Sound on Auditory Localization in Human-Robot Collaboration [20]	Ava	Objective & subjective	Yes	Between sounds, including none	No	In-person; experimental; within-subjects
2018	Multimodal Expression of Artificial Emotion in Social Robots Using Color, Motion and Sound [82]	Custom tabletop robot	Subjective	Yes	Between sounds	Yes	In-person; computer-based; within-subjects
2018	Perception of Mechanical Sounds Inherent to Expressive Gestures of a NAO Robot - Implications for Movement Sonification of Humanoids [36]	NAO	Subjective	Yes	Between sounds	Yes	Online survey; video-based AND audio-only; within-subjects
2018	The Sound or Silence: Investigating the Influence of Robot Noise on Proxemics [153]	Baxter	Objective & subjective	Yes	Between sounds, including none	No	In-person; experimental; between-subjects

Year	**Title**	**Robot**	**Measures**	**Statistical Testing**	**Sound Comparison**	**Multimodal Comparison**	**Study Design**
2019	Augmenting the audio-based expression modality of a non-affective robot [34]	Soft arm	Subjective	Yes	Presence of sound	No	In-person; experimental; between-subjects; group
2019	Establishing Human-Robot Trust through Music-Driven Robotic Emotion Prosody and Gesture [132]	Shimi	Subjective	Yes	Presence of sound	No	In-person; experimental; mixed design
2019	Evaluating the Emotional Valence of Affective Sounds for Child-Robot Interaction [124]	NAO	Subjective	Yes	Between sounds	No	In-person; experimental; mixed design
2019	Impression Change on Nonverbal Non-Humanoid Robot by Interaction with Humanoid Robot [156]	Roomba, NAO	Subjective	Yes	Presence of sound	No	In-person; experimental; within-subjects
2019	Personalized Synthesis of Intentional and Emotional Non-Verbal Sounds for Social Robots [121]	BarBot	Subjective	Yes	Between sounds	No	In-person; experimental; within-subjects
2020	Conveying Robot State and Intent Nonverbally in Military-Relevant Situations: An Exploratory Survey [11]	Jackal	Subjective	No	Between sounds	Yes	In-person; experimental; within-subjects
2020	Correlation Analysis for Predictive Models of Robot User's Impression: A Study on Visual Medium and Mechanical Noise [59]	NAO	Subjective	Yes	Presence of sound	No	In-person; experimental; OR online; video-based; between-subjects

Year	Title	Robot	Measures	Statistical Testing	Sound Comparison	Multimodal Comparison	Study Design
2020	Effects of Auditory Cues on Human-Robot Collaboration [103]	Robot arm	Subjective	Yes	Presence of sound	No	In-person; experimental; within-subjects
2020	Exploring emotion perception in sonic HRI [77]	Pepper	Subjective	Yes	Between sounds	Yes	In-person; experimental; within-subjects
2020	Singing Robots: How Embodiment Affects Emotional Responses to Non-Linguistic Utterances [169]	ROVER	Subjective	Yes	Between sounds	Yes	In-person; experimental; within-subjects
2021	Bringing WALL-E out of the Silver Screen: Understanding How Transformative Robot Sound Affects Human Perception [180]	Cozmo, NAO, TurtleBot, UR5e, Baxter	Subjective	Yes	Presence of sound	Yes	Online survey; video-based; within-subjects
2021	Comparing a Robotic Storyteller versus Audio Book with Integration of Sound Effects and Background Music [143]	NAO	Subjective	Yes	Presence of sound	Yes	Online survey; video-based; between-subjects
2021	Emotion Musical Prosody for Robotic Groups and Entitativity [131]	xArm	Subjective	Yes	Between sounds, including none	No	Online survey; video-based; between-subjects
2021	Exploring Consequential Robot Sound: Should We Make Robots Quiet and Kawaii-et? [178]	UR5e	Subjective	Yes	Between sounds	Yes	Online survey; video-based; within-subjects

Year	**Title**	**Robot**	**Measures**	**Statistical Testing**	**Sound Comparison**	**Multimodal Comparison**	**Study Design**
2022	Perceptual Evaluation of Blended Sonification of Mechanical Robot Sounds Produced by Emotionally Expressive Gestures: Augmenting Consequential Sounds to Improve Non-verbal Robot Communication [35]	NAO	Subjective	Yes	Between sounds, including none	No	In-person; experimental; OR online; video-based; within-subjects

Table 3. Articles that contained human-subjects studies that included created sound as an independent variable, sorted first by year of publication and second by title.

The systematic review contained a total 45 articles with studies on created robot sound. Of the 45 articles, a majority held studies in-person (76%) using robots in a laboratory experiment setting (60%). A variety of robots have been used in the studies, including humanoid robots such as NAO and Pepper, robot arms such as the UR5e, wheeled mobile robots such as Ava and TurtleBot, tabletop robots such as Cozmo, animal robots such as AIBO, and a custom soft robot arm. NAO was the most popular robot and was used in 29% of studies. Study designs mainly compared the presence of sound with no sound conditions (in 24% of articles), different sound designs (47%), or both presence and different designs (27%); two (4%) evaluated sounds in a case study-like way rather than a comparative design. Within-subjects study designs were employed in 76% of the articles, between-subjects designs 22%, and mixed methods designs 9%. Most articles (84%) have applied statistical tests to their measures, which have mainly been subjective (in 82% of articles). Some articles have employed both subjective and objective measures (18%) in their studies, while a few articles have used exclusively objective measures (7%).

Measures have varied greatly between studies. Subjective measures have frequently included valence and energeticness (arousal) from the circumplex model of affect [125], measured using unvalidated questionnaires, AffectButton [14], and the self-assessment manikin (SAM) [12]; general social attributes surveys such as the Godspeed survey, which measures anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety [5], as well as the Robotic Social Attributes Scale (RoSAS), which measures social warmth, competence, and how discomforting a robot is [19]; association with emotions, such as happy, sad, and angry; and preference, usually between sounds or between sound and no sound conditions. Participant-focused measures included the NASA Task Load Index (NASA-TLX) [49], mood, stress, in addition to experience with science, technology, engineering, and mathematics (STEM), electronics, computers, robots in the media, and music. Objective measures have focused around task accuracy and duration. In some studies focused on non-task interactions, reaction time and response type to the robot (e.g., offering to help) were recorded.

Articles have made studied sounds available through multimedia attachments, embedded sound files, and web repositories. However, not all studies made the robot sounds available, and some sounds have become unavailable due to link rot. The incomplete availability of studied sounds has consequences for the longevity of findings in nonverbal robot sound creation. Similarly, not all articles explained sound recording and playback methods, which may create differences, even with identical sound files. We recommend including sounds with archival methods such as multimedia attachments, carefully considering the frequency response of recording and playback methods, and reporting the recording and playback methods used.

4 DISCUSSION

Our exploration of the state and trajectory of the field revealed opportunities in nonverbal robot sound, especially nonverbal robot sound creation. Growth in recent research helps to assert the usefulness and potential of nonverbal robot sound, but many questions remain. For example, how can sound perception be integrated into multimodal perception systems? How can new sounds be integrated into sound source recognition systems? What makes a good consequential robot sound? Transformative? Emotional? Functional? What improvements can be made to robot musicians? Below, we identify research questions of particular interest to us that have been relatively lightly investigated, even within this young and growing research field.

4.1 Paths for Future Research

Consequential robot sound is a ubiquitous feature of human-robot interaction, but few of the reviewed articles investigated consequential sound. These investigations face additional difficulties, as consequential sound inherently corresponds the physical design of the robot, which may be difficult to change, as well as confounding variables like the speed of motion. While changing consequential robot sound requires significant engineering effort, answering some of the many questions revolving around consequential robot sound may benefit all robots.

Open systems and tools may also provide widespread benefits to the nonverbal robot sound field. The reviewed articles included several tools: HARK, which was made open source and is currently maintained [63, 106, 172]), AUROS, which was made open source but is no longer maintained [38, 154], and other tools that were not made publicly available. Common tools can enable more complex investigations of sound, such as through replication and extension of prior work. For instance, in sound creation, sound synthesizers can differ in output sounds even when using similar techniques. We encourage the development of open systems and tools as research contributions such as in the case of our emergent SonifyIt work [179].

In a similar vein, interdisciplinary collaboration with professional musicians and sound designers may improve the complexity and quality of robot sound creation. Work towards improving understanding of the collaborative design process through experience or experimentation may provide valuable research contributions. Mature methods for successful collaboration could allow for broader implementation of research findings in nonverbal robot sound creation on robots, as most roboticists do not also have expertise in sound design.

Lastly, studies on nonverbal robot sound creation often have weaknesses in external validity common to short-term studies in controlled environments. Novelty effects may confound results in short-term robot sound studies, and in real-world environments, robot sound must compete with ambient sounds and travel through different acoustic environments. More in-the-wild and longitudinal study designs can strengthen our understanding of how robot sounds perform in the real world.

4.2 Strengths & Limitations of This Review

This systematic review aggregates more than a quarter century of research, synthesizing 148 articles into new taxonomies for nonverbal robot sound. The strengths of this work include the updated taxonomies of form and function, which carefully integrate prior taxonomies found in the reviewed articles and adjacent fields with best practice scientific communication principles so that researchers from the many disciplines that contribute to nonverbal robot sound may more easily find and read relevant articles. Furthermore, the analysis of study methods complements the analysis and findings in prior work for an updated and broader understanding of nonverbal robot sound creation, in addition to revealing gaps in where and how robot sound has historically been deployed and studied.

We also note the limitations of this work, firstly that the review process did not capture all relevant articles, including recent and highly relevant papers such as by Zahray et al. [177] and Robinson et al. [123]. We hypothesize that the currently lack of common terms and multi-step review process, which necessarily reduced the search space from thousands to hundreds of articles, led to these omissions. While the absence of the aforementioned works do not change our findings, we include them here for the readers' benefit. A second limitation is that the value of introducing new taxonomies and terms rests on how widespread the taxonomies and terms become; rarely-used terms may also become jargon.

4.3 Conclusion

In this article, we proposed new taxonomies for the form and function of sound in human-robot interaction based on the results of a systematic review. These taxonomies will improve the accessibility of the field, making it easier for researchers to share and find related literature. An updated survey of study methods for robot sound creation also reveals opportunities for future work. We highly recommend that researchers in nonverbal robot sound use our updated taxonomies to make nonverbal sound more visible as an important mode of human-robot interaction.

REFERENCES

[1] Roxana Agrigoroaie and Adriana Tapus. 2017. Influence of Robot's Interaction Style on Performance in a Stroop Task. In *Social Robotics*, Abderrahmane Kheddar, Eiichi Yoshida, Shuzhi Sam Ge, Kenji Suzuki, John-John Cabibihan, Friederike Eyssel, and Hongsheng He (Eds.). Vol. 10652. Springer, Cham, Switzerland, 95–104. http://link.springer.com/10.1007/978-3-319-70022-9_10

[2] Takaya Araki, Tomoaki Nakamura, Takayuki Nagai, Kotaro Funakoshi, Mikio Nakano, and Naoto Iwahashi. 2012. Online Object Categorization Using Multimodal Information Autonomously Acquired by a Mobile Robot. *Advanced Robotics* 26, 17 (Dec. 2012), 1995–2020. https://doi.org/10.1080/01691864.2012.728693

[3] Ahmad Faris Azmin, Syamimi Shamsuddin, and Hanafiah Yussof. 2016. HRI observation with My Keepon robot using Kansei Engineering approach. In *Proceedings of the IEEE International Symposium on Robotics and Manufacturing Automation (ROMA)*. IEEE, Ipoh, Malaysia, 1–6. https://doi.org/10.1109/ROMA.2016.7847831

[4] Jaclyn A. Barnes, Chung Hyuk Park, Ayanna Howard, and Myounghoon Jeon. 2021. Child-Robot Interaction in a Musical Dance Game: An Exploratory Comparison Study between Typically Developing Children and Children with Autism. *International Journal of Human–Computer Interaction* 37, 3 (Feb. 2021), 249–266. https://doi.org/10.1080/10447318.2020.1819667

[5] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1, 1 (Jan. 2009), 71—-81. https://doi.org/10.1007/s12369-008-0001-3

[6] Jon Bellona, Lin Bai, Luke Dahl, and Amy LaViers. 2017. Empirically Informed Sound Synthesis Application for Enhancing the Perception of Expressive Robotic Movement. In *Proceedings of the International Conference on Auditory Display (ICAD)*. The International Community for Auditory Display, Happy Valley, PA, USA, 73–80. https://doi.org/10.21785/icad2017.049

[7] Maren Bennewitz, Felix Faber, Dominik Joho, Michael Schreiber, and Sven Behnke. 2005. Towards a humanoid museum guide robot that interacts with multiple persons. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*. IEEE, San Diego, CA, USA, 418–423. https://doi.org/10.1109/ICHR.2005.1573603

[8] Laura Boccanfuso, Elizabeth S. Kim, James C. Snider, Quan Wang, Carla A. Wall, Lauren DiNicola, Gabriella Greco, Lilli Flink, Sharlene Lansiquot, Pamela Ventola, Katarzyna Chawarska, Brian Scassellati, and Frederick Shic. 2015. Autonomously detecting interaction with an affective robot to explore connection to developmental ability. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, Xi'an, China, 1–7. https://doi.org/10.1109/ACII.2015.7344543

[9] Gabriele Bolano, Arne Roennau, and Ruediger Dillmann. 2018. Transparent Robot Behavior by Adding Intuitive Visual and Acoustic Feedback to Motion Replanning. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Nanjing, China, 1075–1080. https://doi.org/10.1109/ROMAN.2018.8525671

[10] Oliver Bones, Trevor J. Cox, and William J. Davies. 2018. Sound Categories: Category Formation and Evidence-Based Taxonomies. *Frontiers in Psychology* 9 (2018), 1–17. https://www.frontiersin.org/article/10.3389/fpsyg.2018.01277

[11] Lauren Boos and Lilia Moshkina. 2020. Conveying Robot State and Intent Nonverbally in Military-Relevant Situations: An Exploratory Survey. In *Proceedings of the International Conference on Applied Human Factors and Ergonomics (AHFE)*, Jessie Chen (Ed.), Vol. 962. Springer, San Diego, CA, USA, 181–193. https://doi.org/10.1007/978-3-030-20467-9_17

[12] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (March 1994), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

[13] Derek Brock and Eric Martinson. 2006. Exploring the utility of giving robots auditory perspective-taking abilities. In *Proceedings of the International Conference on Auditory Display (ICAD)*. The International Community for Auditory Display, London, United Kingdom, 250–253. https://www.icad.org/Proceedings/2006/BrockMartinson2006.pdf

[14] Joost Broekens and Willem-Paul Brinkman. 2013. AffectButton: A method for reliable and valid affective self-report. *International Journal of Human-Computer Studies* 71, 6 (2013), 641–667. https://doi.org/10.1016/j.ijhcs.2013.02.003 Place: Netherlands Publisher: Elsevier Science.

[15] Rodney A. Brooks, Cynthia Breazeal, Matthew Marjanović, Brian Scassellati, and Matthew M. Williamson. 1998. The Cog Project: Building a Humanoid Robot. In *Proceedings of the International Workshop on Computation for Metaphors, Analogy, and Agents (CMAA) (Lecture Notes in Computer Science, Vol. 1562)*, G. Goos, J. Hartmanis, J. van Leeuwen, and Chrystopher L. Nehaniv (Eds.). Springer, Berlin, Heidelberg, 52–87. https://doi.org/10.1007/3-540-48834-0_5

[16] Sinan Bökesoy. 2014. A Recursive Mapping System For Motion And Sound In A Robot Between Human Interaction Design. In *Proceedings of the Sound and Music Computing Conference (SMC) and International Computer Music Conference (ICMC)*. Sound and Music Computing Network, Athens, Greece, 1649–1654. https://doi.org/10.5281/zenodo.850949

[17] Lech Błażejewski. 2005. Spatial Sound Localization for Humanoid. In *Monitoring, Security, and Rescue Techniques in Multiagent Systems*. AINSC, Vol. 28. Springer, Berlin/Heidelberg, 527–537. http://link.springer.com/10.1007/3-540-32370-8_42

[18] Antonio Camurri, Matteo Ricchetti, and Riccardo Trocca. 1999. EyesWeb-toward gesture and affect recognition in dance/music interactive systems. In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems (ICMC)*, Vol. 1. IEEE, Florence, Italy, 643–648. https://doi.org/10.1109/MMCS.1999.779275

[19] Colleen M. Carpinella, Alisa B. Wyman, Michael A. Perez, and Steven J. Stroessner. 2017. The Robotic Social Attributes Scale (RoSAS): Development and Validation. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Vienna, Austria, 254–262. https://ieeexplore.ieee.org/document/8534914

[20] Elizabeth Cha, Naomi T. Fitter, Yunkyung Kim, Terrence Fong, and Maja J. Matarić. 2018. Effects of Robot Sound on Auditory Localization in Human-Robot Collaboration. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Chicago, IL, USA, 434–442. https://doi.org/10.1145/3171221.3171285

[21] Elizabeth Cha, Maja Mataric, and Terrence Fong. 2016. Nonverbal signaling for non-humanoid robots during human-robot collaboration. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Christchurch, New Zealand, 601–602. https://doi.org/10.1109/HRI.2016.7451876

[22] Elizabeth Cha and Maja Matarić. 2016. Using nonverbal signals to request help during human-robot collaboration. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE/RSJ, Daejeon, South Korea, 5070–5076. https://doi.org/10.1109/IROS.2016.7759744

[23] Sutirtha Chakraborty and Joseph Timoney. 2020. Robot Human Synchronization for Musical Ensemble: Progress and Challenges. In *Proceedings of the International Conference on Robotics and Automation Engineering (ICRAE)*. IEEE, Singapore, Singapore, 93–99. https://doi.org/10.1109/ICRAE50850.2020.9310916

[24] Kyoung Soo Chun, Eun-Sook Jee, and Dong-Soo Kwon. 2012. Novel musical notation for emotional sound expression of interactive robot. In *Proceedings of the International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*. IEEE, Daejeon, Korea (South), 88–89. https://doi.org/10.1109/URAI.2012.6462939

[25] F. Ciardo, D. De Tommaso, and A. Wykowska. 2021. Effects of erring behavior in a human-robot joint musical task on adopting Intentional Stance toward the iCub robot. In *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Vancouver, BC, Canada, 698–703. https://doi.org/10.1109/RO-MAN50785.2021.9515434

[26] Luke Dahl, Jon Bellona, Lin Bai, and Amy LaViers. 2017. Data-Driven Design of Sound for Enhancing the Perception of Expressive Robotic Movement. In *Proceedings of the International Conference on Movement Computing (MOCO)*. ACM, London, United Kingdom, 1–8. https://doi.org/10.1145/3077981.3078047

[27] Itamar Eliakim, Zahi Cohen, Gabor Kosa, and Yossi Yovel. 2018. A fully autonomous terrestrial bat-like acoustic robot. *PLOS Computational Biology* 14, 9 (Sept. 2018), e1006406.

[28] Jing Fan, Akshith Ullal, Linda Beuscher, Lorraine C. Mion, Paul Newhouse, and Nilanjan Sarkar. 2021. Field Testing of Ro-Tri, a Robot-Mediated Triadic Interaction for Older Adults. *International Journal of Social Robotics* 13, 7 (Nov. 2021), 1711–1727. https://doi.org/10.1007/s12369-021-00760-2

[29] Javier Fernandez De Gorostiza Luengo, Fernando Alonso Martin, Alvaro Castro-Gonzalez, and Miguel Angel Salichs. 2017. Sound Synthesis for Communicating Nonverbal Expressive Cues. *IEEE Access* 5 (2017), 1941–1957. https://doi.org/10.1109/ACCESS.2017.2658726

[30] Kerstin Fischer, Lars Christian Jensen, and Leon Bodenhagen. 2014. To Beep or Not to Beep Is Not the Whole Question. In *Social Robotics*, Michael Beetz, Benjamin Johnston, and Mary-Anne Williams (Eds.). Vol. 8755. Springer, Cham, Switzerland, 156–165. http://link.springer.com/10.1007/978-3-319-11973-1_16

[31] Kerstin Fischer, Bianca Soto, Caroline Pantofaru, and Leila Takayama. 2014. Initiating interactions in order to get help: Effects of social framing on people's responses to robots' requests for assistance. In *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Edinburgh, UK, 999–1005. https://doi.org/10.1109/ROMAN.2014.6926383

[32] Evelyn Florentine, Mark Adam Ang, Scott Drew Pendleton, Hans Andersen, and Marcelo H. Ang. 2016. Pedestrian Notification Methods in Autonomous Vehicles for Multi-Class Mobility-on-Demand Service. In *Proceedings of the International Conference on Human Agent Interaction (HAI)*. ACM, Singapore, Singapore, 387–392. https://doi.org/10.1145/2974804.2974833

[33] Stavroula-Evita Fotinea, Athanasia-Lida Dimou, Eleni Efthimiou, and Theodore Goulas. 2014. The annotation scheme of the MOBOT dataset. In *Proceedings of the IEEE Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, Vietri sul Mare, Italy, 251–256. https://doi.org/10.1109/CogInfoCom.2014.7020456

[34] Morten Roed Frederiksen and Kasper Stoey. 2019. Augmenting the audio-based expression modality of a non-affective robot. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, Cambridge, United Kingdom, 144–149. https://doi.org/10.1109/ACII.2019.8925510

[35] Emma Frid and Roberto Bresin. 2022. Perceptual Evaluation of Blended Sonification of Mechanical Robot Sounds Produced by Emotionally Expressive Gestures: Augmenting Consequential Sounds to Improve Non-verbal Robot Communication. *International Journal of Social Robotics* 14, 2 (March 2022), 357–372. https://doi.org/10.1007/s12369-021-00788-4

[36] Emma Frid, Roberto Bresin, and Simon Alexandersson. 2018. Perception Of Mechanical Sounds Inherent To Expressive Gestures Of A Nao Robot - Implications For Movement Sonification Of Humanoids. In *Proceedings of the Sound and Music Computing Conference (SMC)*. Sound and Music Computing Network, Limassol, Cyprus, 43–51. https://doi.org/10.5281/zenodo.1422499

[37] David Gerhard. 2003. *Audio Signal Classification: History and Current Techniques*. Technical Report TR-CS 2003-07. University of Regina, Regina, Saskatchewan, Canada. https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.11.778

[38] Theodoros Giannakopoulos and Georgios Siantikos. 2016. A ROS framework for audio-based activity recognition. In *Proceedings of the ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA)*. ACM, Corfu Island, Greece, 1–4. https://doi.org/10.1145/2910674.2935858

[39] Naomi Gildert, Alan G. Millard, Andrew Pomfret, and Jon Timmis. 2018. The Need for Combining Implicit and Explicit Communication in Cooperative Robotic Systems. *Frontiers in Robotics and AI* 5 (2018), 1–6. https://www.frontiersin.org/article/10.3389/frobt.2018.00065

[40] Google. 2022. *Google Books Ngram Viewer.* Google. Retrieved 2022-06-01 from https://books.google.com/ngrams/ Search parameters: terms: vocable,non-linguistic utterance,semantic-free utterance,paralinguistic utterance,gibberish speech; period: 1996 to 2019; database: English (2019); case-insensitive; smoothing of 0..

[41] Google. 2022. *Google Books Ngram Viewer.* Google. Retrieved 2022-06-01 from https://books.google.com/ngrams/ Search parameters: terms: self-noise,ego-noise; period: 1996 to 2019; database: English (2019); case-insensitive; smoothing of 0..

[42] Google. 2022. *Google Search Trends.* Google. Retrieved 2022-06-01 from https://trends.google.com/trends/ Search parameters: terms: vocable,non-linguistic utterance,semantic-free utterance,paralinguistic utterance,gibberish speech; database: Worldwide; period: 2004 - present; All categories; Web Search..

[43] Google. 2022. *Google Search Trends.* Google. Retrieved 2022-06-01 from https://trends.google.com/trends/ Search parameters: terms: self-noise,ego-noise; database: Worldwide; period: 2004 - present; All categories; Web Search..

[44] Javi F. Gorostiza, Ramon Barber, Alaa M. Khamis, Maria Malfaz, Rakel Pacheco, Rafael Rivas, Ana Corrales, Elena Delgado, and Miguel A. Salichs. 2006. Multimodal Human-Robot Interaction Framework for a Personal Robot. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Hatfield, United Kingdom, 39–44. https://doi.org/10.1109/ROMAN.2006.314392

[45] Lacrimioara Grama and Corneliu Rusu. 2018. Adding audio capabilities to TIAGo service robot. In *Proceedings of the International Symposium on Electronics and Telecommunications (ISETC)*. IEEE, Timisoara, Romania, 1–4. https://doi.org/10.1109/ISETC.2018.8583897

[46] Ellen C. Haas. 2007. Integrating Auditory Warnings with Tactile Cues in Multimodal Displays for Challenging Environments. In *Proceedings of the International Conference on Auditory Display (ICAD)*. The International Community for Auditory Display, Montreal, Canada, 126–130. https://doi.org/1853/50004

[47] Andrew H. Hales, Kipling D. Williams, and and Joel Rector. 2017. Alienating the Audience: How Abbreviations Hamper Scientific Communication. *APS Observer* 30 (Jan. 2017), 1–6. https://www.psychologicalscience.org/observer/alienating-the-audience-how-abbreviations-hamper-scientific-communication

[48] W. K. N. Hansika, Lakindu Yasassri Nanayakkara, Adhisha Gammanpila, and Ravindra de Silva. 2020. AuDimo: A Musical Companion Robot to Switching Audio Tracks by Recognizing the Users Engagement. In *Proceedings of the International Conference on Human-Computer Interaction (HCII) - Late Breaking Papers: Multimodality and Intelligence*, Constantine Stephanidis, Masaaki Kurosu, Helmut Degen, and Lauren Reinerman-Jones (Eds.), Vol. 12424. Springer, Copenhagen, Denmark, 89–106. https://doi.org/10.1007/978-3-030-60117-1_7

[49] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, Peter A. Hancock and Najmedin Meshkati (Eds.). Human Mental Workload, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[50] Helen Hastie, Pasquale Dente, Dennis Küster, and Arvid Kappas. 2016. Sound emblems for affective multimodal output of a robotic tutor: a perception study. In *Proceedings of the ACM International Conference on Multimodal Interaction (ICMI)*. ACM, Tokyo, Japan, 256–260. https://doi.org/10.1145/2993148.2993169

[51] Weipeng He, Petr Motlicek, and Jean-Marc Odobez. 2018. Deep Neural Networks for Multiple Speaker Detection and Localization. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, Brisbane, QLD, Australia, 74–79. https://doi.org/10.1109/ICRA.2018.8461267

[52] Thomas Hermann. 2008. Taxonomy and Definitions for Sonification and Auditory Display. In *Proceedings of the International Conference on Auditory Display (ICAD)*. International Community for Auditory Display, Paris, France, 1–8.

[53] Thomas Hermann, Christian Niehus, and Helge Ritter. 2003. Interactive visualization and sonification for monitoring complex processes. In *Proceedings of the International Conference on Auditory Display (ICAD)*. The International Community for Auditory Display, Boston, MA, USA, 247–250. https://doi.org/1853/50467

[54] Guy Hoffman. 2012. Dumb robots, smart phones: A case study of music listening companionship. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Paris, France, 358–363. https://doi.org/10.1109/ROMAN.2012.6343779

[55] Guy Hoffman, Shira Bauman, and Keinan Vanunu. 2016. Robotic experience companionship in music listening and video watching. *Personal and Ubiquitous Computing* 20, 1 (Feb. 2016), 51–63. https://doi.org/10.1007/s00779-015-0897-1

[56] Guy Hoffman and Keinan Vanunu. 2013. Effects of robotic companionship on music enjoyment and agent perception. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Tokyo, Japan, 317–324. https://doi.org/10.1109/HRI.2013.6483605

[57] Gökhan Ince, Taha Berkay Duman, Rabia Yorganci, and Hatice Kose. 2015. Towards a robust drum stroke recognition system for human robot interaction. In *Proceedings of the IEEE/SICE International Symposium on System Integration (SII)*. IEEE/RSJ, Nagoya, Japan, 744–749. https://doi.org/10.1109/SII.2015.7405072

[58] Gökhan Ince, Rabia Yorganci, Ahmet Ozkul, Taha Berkay Duman, and Hatice Köse. 2021. An audiovisual interface-based drumming system for multimodal human–robot interaction. *Journal on Multimodal User Interfaces* 15, 4 (Dec. 2021), 413–428. https://doi.org/10.1007/s12193-020-00352-w

[59] Takamune Izui and Gentiane Venture. 2020. Correlation Analysis for Predictive Models of Robot User's Impression: A Study on Visual Medium and Mechanical Noise. *International Journal of Social Robotics* 12, 2 (May 2020), 425–439. https://doi.org/10.1007/s12369-019-00601-3

[60] Shruti Jaiswal, Ayush Jain, and G.C. Nandi. 2020. Image based Emotional State Prediction from Multiparty Audio Conversation. In *Proceedings of the IEEE Pune Section International Conference (PuneCon)*. IEEE, Pune, India, 77–82. https://doi.org/10.1109/PuneCon50868.2020.9362475

[61] Maxime Janvier, Xavier Alameda-Pineda, Laurent Girin, and Radu Horaud. 2012. Sound-event recognition with a companion humanoid. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids)*. IEEE, Osaka, Japan, 104–111. https://doi.org/10.1109/HUMANOIDS.2012.6651506

[62] Maxime Janvier, Xavier Alameda-Pineda, Laurent Girin, and Radu Horaud. 2014. Sound representation and classification benchmark for domestic robots. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, Hong Kong, China, 6285–6292. https://doi.org/10.1109/ICRA.2014.6907786

[63] Honda Research Institute Japan. 2022. *Honda Research Institute Japan Audition for Robots with Kyoto University (HARK)*. Honda Research Institute Japan. Retrieved 2022-06-01 from https://hark.jp/

[64] Eunju Jeong, Gyu Hyun Kwon, and Junseop So. 2017. Exploring the taxonomic and associative link between emotion and function for robot sound design. In *Proceedings of the International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*. IEEE, Jeju, South Korea, 641–643. https://doi.org/10.1109/URAI.2017.7992692

[65] Pengju Jin, Eshed Ohn-Bar, Kris Kitani, and Chieko Asakawa. 2019. A-EXP4: Online Social Policy Learning for Adaptive Robot-Pedestrian Interaction. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE/RSJ, Macau, China, 5086–5093. https://doi.org/10.1109/IROS40897.2019.8967737

[66] Gunnar Johannsen. 2002. Auditory display of directions and states for mobile systems. In *Proceedings of the International Conference on Auditory Display (ICAD)*. The International Community for Auditory Display, Kyoto, Japan, 1–6. https://doi.org/1853/51337

[67] Michiel Joosse, Manja Lohse, and Vanessa Evers. 2014. Sound over matter: the effects of functional noise, robot size and approach velocity in human-robot encounters. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Bielefeld, Germany, 184–185. https://doi.org/10.1145/2559636.2559822

[68] Hyun-Don Kim, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno. 2009. Human Tracking System Integrating Sound and Face Localization Using an Expectation-Maximization Algorithm in Real Environments. *Advanced Robotics* 23, 6 (Jan. 2009), 629–653. https://doi.org/10.1163/156855309X431659

[69] Heather Knight. 2011. Eight Lessons Learned about Non-verbal Interactions through Robot Theater. In *Social Robotics*, Bilge Mutlu, Christoph Bartneck, Jaap Ham, Vanessa Evers, and Takayuki Kanda (Eds.). Vol. 7072. Springer, Heidelberg, Germany, 42–51. http://link.springer.com/10.1007/978-3-642-25504-5_5

[70] Kheng Lee Koay, Michael L. Walters, Alex May, Anna Dumitriu, Bruce Christianson, Nathan Burke, and Kerstin Dautenhahn. 2013. Exploring Robot Etiquette: Refining a HRI Home Companion Scenario Based on Feedback from Two Artists Who Lived with Robots in the UH Robot House. In *Social Robotics*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Guido Herrmann, Martin J. Pearson, Alexander Lenz, Paul Bremner, Adam Spiers, and Ute Leonards (Eds.). Vol. 8239. Springer, Cham, Switzerland, 290–300. http://link.springer.com/10.1007/978-3-319-02675-6_29

[71] Kazuki Kobayashi, Yasuhiko Kitamura, and Seiji Yamada. 2007. Action Sloping as a Way for Users to Notice a Robot's Function. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Jeju, South Korea, 445–450. https://doi.org/10.1109/ROMAN.2007.4415125

[72] Christian Kroos, Damith C. Herath, and Stelarc. 2010. The Articulated Head pays attention. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Osaka, Japan, 357–357. https://doi.org/10.1109/HRI.2010.5453164

[73] Michael Krzyżaniak. 2021. Musical robot swarms, timing, and equilibria. *Journal of New Music Research* 50, 3 (May 2021), 279–297. https://doi.org/10.1080/09298215.2021.1910313

[74] Hoshito Kudo, Tomoya Koizumi, Tsuyoshi Nakamura, Masayoshi Kanoh, and Koji Yamada. 2016. Behavior Model for Hearing-Dog Robot. In *Proceedings of the Joint International Conference on Soft Computing and Intelligent Systems (SCIS) and International Symposium on Advanced Intelligent Systems (ISIS)*. IEEE, Sapporo, Japan, 260–265. https://doi.org/10.1109/SCIS-ISIS.2016.0063

[75] Shan Lakhmani, Julian Abich, Daniel Barber, and Jessie Chen. 2016. A Proposed Approach for Determining the Influence of Multimodal Robot-of-Human Transparency Information on Human-Agent Teams. In *Proceedings of the International Conference on Augmented Cognition (AC)*, Dylan D. Schmorrow and Cali M. Fidopiastis (Eds.). Springer, Toronto, Canada, 296–307. https://doi.org/10.1007/978-3-319-39952-2_29

[76] Lau Langeveld, René van Egmond, Reinier Jansen, and Elif Özcan. 2013. Product Sound Design: Intentional and Consequential Sounds. *Advances in Industrial Design Engineering* (March 2013). https://doi.org/10.5772/55274

[77] Adrian B. Latupeirissa, Claudio Panariello, and Roberto Bresin. 2020. Exploring emotion perception in sonic HRI. In *Proceedings of the Sound and Music Computing Conference (SMC)*. Sound and Music Computing Network. https://doi.org/10.5281/ZENODO.3898928

[78] Jeannie S. Lee, Muhamed Fauzi Bin Abbas, Chee Kiat Seow, Qi Cao, Kar Peo Yar, Sye Loong Keoh, and Ian McLoughlin. 2021. Non-Verbal Auditory Aspects of Human-Service Robot Interaction. In *Proceedings of the IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*. IEEE, Singapore, Singapore, 1–5. https://doi.org/10.1109/SOLI54607.2021.9672366

[79] Angelica Lim, Takeshi Mizumoto, Tetsuya Ogata, and Hiroshi G. Okuno. 2012. A Musical Robot that Synchronizes with a Coplayer Using Non-Verbal Cues. *Advanced Robotics* 26, 3-4 (Jan. 2012), 363–381. https://doi.org/10.1163/156855311X614626

[80] Dilip Kumar Limbu, Wong Chern Yuen Anthony, Tay Hwang Jian Adrian, Tran Anh Dung, Tan Yeow Kee, Tran Huy Dat, Wong Hong Yee Alvin, Ng Wen Zheng Terence, Jiang Ridong, and Li Jun. 2013. Affective social interaction with CuDDler robot. In *Proceedings of the IEEE Conference on Robotics, Automation and Mechatronics (RAM)*. IEEE, Manila, Philippines, 179–184. https://doi.org/10.1109/RAM.2013.6758580

[81] Yong (Yates) Lin, Zhengyi Le, Eric Becker, and Fillia Makedon. 2010. Acoustical implicit communication in human-robot interaction. In *Proceedings of the ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA)*. ACM, Samos, Greece, 1–5. https://doi.org/10.1145/1839294.1839300

[82] Diana Löffler, Nina Schmidt, and Robert Tscharn. 2018. Multimodal Expression of Artificial Emotion in Social Robots Using Color, Motion and Sound. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Chicago, IL, USA, 334–343. https://doi.org/10.1145/3171221.3171261

[83] Marc-Antoine Maheux, Charles Caya, Dominic Létourneau, and François Michaud. 2022. T-Top, a SAR Experimental Platform. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Sapporo, Hokkaido, Japan, 904–908. https://doi.org/10.5555/3523760.3523902

[84] Umberto Maniscalco, Pietro Storniolo, and Antonio Messina. 2022. Bidirectional Multi-modal Signs of Checking Human-Robot Engagement and Interaction. *International Journal of Social Robotics* (April 2022). https://doi.org/10.1007/s12369-021-00855-w

[85] Eric Martinson and Veera Ganesh Yalla. 2014. Guiding computational perception through a shared auditory space. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE/RSJ, Chicago, IL, USA, 3156–3161. https://doi.org/10.1109/IROS.2014.6942999

[86] Lourdes Martínez-Villaseñor and Hiram Ponce. 2019. A concise review on sensor signal acquisition and transformation applied to human activity recognition and human–robot interaction. *International Journal of Distributed Sensor Networks* 15, 6 (June 2019). https://doi.org/10.1177/1550147719853987

[87] Naoki Masuyama, Chee Seng Chan, Naoyuki Kuobota, and Jinseok Woo. 2012. Computational Intelligence for Human Interactive Communication of Robot Partners. In *Proceedings of the Pacific Rim International Conference on Artificial Intelligence (PRICAI): Trends in Artificial Intelligence*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Patricia Anthony, Mitsuru Ishizuka, and Dickson Lukose (Eds.), Vol. 7458. Springer, Berlin, Heidelberg, 771–776. https://doi.org/10.1007/978-3-642-32695-0_71

[88] Louis McCallum and Peter W McOwan. 2014. Shut up and play: A musical approach to engagement and social presence in Human Robot Interaction. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Edinburgh, UK, 949–954. https://doi.org/10.1109/ROMAN.2014.6926375

[89] Louis McCallum and Peter W. McOwan. 2015. Face the Music and Glance: How Nonverbal Behaviour Aids Human Robot Relationships Based in Music. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Portland, OR, USA, 237–244. https://doi.org/10.1145/2696454.2696477

[90] Louis McCallum and Peter W. McOwan. 2018. Extending Human–Robot Relationships Based in Music With Virtual Presence. *IEEE Transactions on Cognitive and Developmental Systems* 10, 4 (Dec. 2018), 955–960. https://doi.org/10.1109/TCDS.2017.2779218

[91] Merriam-Webster. 2022. *noise*. Merriam-Webster. Retrieved 2022-06-01 from https://www.merriam-webster.com/dictionary/noise

[92] Merriam-Webster. 2022. *Nonverbal*. Merriam-Webster. Retrieved 2022-06-01 from https://www.merriam-webster.com/dictionary/nonverbal

[93] Merriam-Webster. 2022. *Sound*. Merriam-Webster. Retrieved 2022-06-01 from https://www.merriam-webster.com/dictionary/sound

[94] Alexander Mertens, Sinem Kuz, Antje Heinicke, Marcel Mayer, Bernhard Kausch, and Christopher Schlick. 2011. User focused design of human-robot interaction for people suffering from unusual ailments. In *Proceedings of the International Conference on Human System Interactions (HSI)*. IEEE, Yokohama, Japan, 44–51. https://doi.org/10.1109/HSI.2011.5937341

[95] Marek P. Michalowski, Selma Sabanovic, and Hideki Kozima. 2007. A Dancing Robot for Rhythmic Social Interaction. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Arlington, VA, USA, 89–96. https://doi.org/10.1145/1228716.1228729

[96] Marek P. Michalowski, Reid Simmons, and Hideki Kozima. 2009. Rhythmic attention in child-robot dance play. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Toyama, Japan, 816–821. https://doi.org/10.1109/ROMAN.2009.5326143

[97] Naeeum Assif Mirza, Chrystopher L. Nehaniv, Kerstin Dautenhahn, and René te Boekhorst. 2008. Developing social action capabilities in a humanoid robot using an interaction history architecture. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*. IEEE, Daejeon, 609–616. https://doi.org/10.1109/ICHR.2008.4756013

[98] Dylan Moore, Hamish Tennent, Nikolas Martelaro, and Wendy Ju. 2017. Making Noise Intentional: A Study of Servo Sound Perception. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Vienna, Austria, 12–21. https://doi.org/10.1145/2909824.3020238

[99] Lorena Muscar, Lacrimioara Grama, and Corneliu Rusu. 2021. Sound Classification by the TIAGo Service Robot for Healthcare Applications. In *Proceedings of the International Symposium on Signals, Circuits and Systems (ISSCS)*. IEEE, Iasi, Romania, 1–4. https://doi.org/10.1109/ISSCS52333.2021.9497410

[100] Neumann. 2022. *What is Self-Noise (or Equivalent Noise Level)?* Neumann. Retrieved 2022-06-01 from https://www.neumann.com/homestudio/en/what-is-self-noise-or-equivalent-noise-level

[101] Anton Nijholt. 2018. Robotic Stand-Up Comedy: State-of-the-Art. In *Distributed, Ambient and Pervasive Interactions: Understanding Humans*, Norbert Streitz and Shin'ichi Konomi (Eds.). Vol. 10921. Springer, Cham, Switzerland, 391–410. http://link.springer.com/10.1007/978-3-319-91125-0_32

[102] R. Nikolaidis and Gil Weinberg. 2010. Playing with the masters: A model for improvisatory musical interaction between robots and humans. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Viareggio, Italy, 712–717. https://doi.org/10.1109/ROMAN.2010.5598621

[103] Jumpei Okimoto and Mihoko Niitsuma. 2020. Effects of Auditory Cues on Human-Robot Collaboration. In *Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE)*. IEEE, Delft, Netherlands, 1572–1577. https://doi.org/10.1109/ISIE45063.2020.9152413

[104] Hiroshi G. Okuno and Kazuhiro Nakadai. 2003. Real-time Sound Source Localization and Separation based on Active Audio-Visual Integration. In *Computational Methods in Neural Modeling*, Gerhard Goos, Juris Hartmanis, Jan van Leeuwen, José Mira, and José R. Álvarez (Eds.). Vol. 2686. Springer, Berlin, Heidelberg, 118–125. http://link.springer.com/10.1007/3-540-44868-3_16

[105] Hiroshi G. Okuno, Kazuhiro Nakadai, Ken-Ichi Hidai, Hiroshi Mizoguchi, and Hiroaki Kitano. 2003. Human–robot non-verbal interaction empowered by real-time auditory and visual multiple-talker tracking. *Advanced Robotics* 17, 2 (Jan. 2003), 115–130. https://doi.org/10.1163/156855303321165088

[106] Hiroshi G. Okuno, Kazuhiro Nakadai, and Hyun-Don Kim. 2011. Robot Audition: Missing Feature Theory Approach and Active Audition. In *Proceedings of the International Symposium on Robotics Research (ISRR)*, Bruno Siciliano, Oussama Khatib, Frans Groen, Cédric Pradalier, Roland Siegwart, and Gerhard Hirzinger (Eds.), Vol. 70. Springer, Lucerne, Switzerland, 227–244. https://doi.org/10.1007/978-3-642-19457-3_14

[107] Hiroshi G. Okuno, Kazuhiro Nakadai, and Hiroaki Kitano. 2002. Social Interaction of Humanoid Robot Based on Audio-Visual Tracking. In *Developments in Applied Artificial Intelligence*, G. Goos, J. Hartmanis, J. van Leeuwen, Tim Hendtlass, and Moonis Ali (Eds.). Vol. 2358. Springer, Berlin, Heidelberg, 725–735. http://link.springer.com/10.1007/3-540-48035-8_70

[108] Hiroshi G. Okuno, Kazuhiro Nakadai, and Hiroaki Kitano. 2003. Design and Implementation of Personality of Humanoids in Human Humanoid Non-verbal Interaction. In *Developments in Applied Artificial Intelligence*, Paul W. H. Chung, Chris Hinde, and Moonis Ali (Eds.). Vol. 2718. Springer, Berlin, Heidelberg, 662–673. http://link.springer.com/10.1007/3-540-45034-3_67

[109] Hiroshi G. Okuno, Kazuhiro Nakadai, and Hiroaki Kitano. 2003. Realizing personality in audio-visually triggered non-verbal behaviors. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Vol. 1. IEEE, Taipei, Taiwan, 392–397. https://doi.org/10.1109/ROBOT.2003.1241627

[110] Joao Lobato Oliveira, Gökhan Ince, Keisuke Nakamura, Kazuhiro Nakadai, Hiroshi G. Okuno, Luis Paulo Reis, and Fabien Gouyon. 2012. An active audition framework for auditory-driven HRI: Application to interactive robot dancing. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Paris, France, 1078–1085. https://doi.org/10.1109/ROMAN.2012.6343892

[111] Jeong-Sik Park, Gil-Jin Jang, and Yong-Ho Seo. 2012. Music-aided affective interaction between human and service robot. *EURASIP Journal on Audio, Speech, and Music Processing* 2012, 1 (Dec. 2012), 5. https://doi.org/10.1186/1687-4722-2012-5

[112] Hannah R. M. Pelikan, Mathias Broth, and Leelo Keevallik. 2020. "Are You Sad, Cozmo?": How Humans Make Sense of a Home Robot's Emotion Displays. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Cambridge United Kingdom, 461–470. https://doi.org/10.1145/3319502.3374814

[113] Hua Peng, Changle Zhou, Huosheng Hu, Fei Chao, and Jing Li. 2015. Robotic Dance in Social Robotics—A Taxonomy. *IEEE Transactions on Human-Machine Systems* 45, 3 (June 2015), 281–293. https://doi.org/10.1109/THMS.2015.2393558

[114] Thales Roel P. Pessanha, Higor Camporez, Jônatas Manzolli, Bruno Sanches Masiero, Leandro Costalonga, and Tiago Fernandes Tavares. 2021. Virtual Robotic Musicianship: Challenges and Opportunities. In *Proceedings of the Sound and Music Computing Conference (SMC)*. Sound and Music Computing Network. https://doi.org/10.5281/zenodo.5040623

[115] Shokoofeh Pourmehr, Valiallah Monajjemi, Jens Wawerla, Richard Vaughan, and Greg Mori. 2013. A robust integrated system for selecting and commanding multiple mobile robots. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, Karlsruhe, Germany, 2874–2879. https://doi.org/10.1109/ICRA.2013.6630975

[116] Robin Read and Tony Belpaeme. 2012. How to use non-linguistic utterances to convey emotion in child-robot interaction. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Boston, MA, USA, 219. https://doi.org/10.1145/2157689.2157764

[117] Robin Read and Tony Belpaeme. 2013. People interpret robotic non-linguistic utterances Categorically. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Tokyo, Japan, 209–210. https://doi.org/10.1109/HRI.2013.6483575

[118] Robin Read and Tony Belpaeme. 2014. Non-linguistic utterances should be used alongside language, rather than on their own or as a replacement. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Bielefeld, Germany, 276–277. https://doi.org/10.1145/2559636.2559836

[119] Robin Read and Tony Belpaeme. 2014. Situational context directs how people affectively interpret robotic non-linguistic utterances. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Bielefeld, Germany, 41–48. https://doi.org/10.1145/2559636.2559680

[120] Robin G. Read and Tony Belpaeme. 2010. Interpreting non-linguistic utterances by robots: studying the influence of physical appearance. In *Proceedings of the International Workshop on Affective Interaction in Natural Environments (AFFINE)*. ACM, Firenze, Italy, 65. https://doi.org/10.1145/1877826.1877843

[121] Hannes Ritschel, Ilhan Aslan, Silvan Mertes, Andreas Seiderer, and Elisabeth Andre. 2019. Personalized Synthesis of Intentional and Emotional Non-Verbal Sounds for Social Robots. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, Cambridge, United Kingdom, 1–7. https://doi.org/10.1109/ACII.2019.8925487

[122] Ben Robins, Kerstin Dautenhahn, and Paul Dickerson. 2009. From Isolation to Communication: A Case Study Evaluation of Robot Assisted Play for Children with Autism with a Minimally Expressive Humanoid Robot. In *Proceedings of the International Conference on Advances in Computer-Human Interactions (ACHI)*. IEEE, Cancun, Mexico, 205–211. https://doi.org/10.1109/ACHI.2009.32

[123] Frederic Anthony Robinson, Mari Velonaki, and Oliver Bown. 2021. Smooth Operator: Tuning Robot Perception Through Artificial Movement Sound. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. Association for Computing Machinery, New York, NY, USA, 53–62. https://doi.org/10.1145/3434073.3444658

[124] Silvia Rossi, Elena Dell'Aquila, and Benedetta Bucci. 2019. Evaluating the Emotional Valence of Affective Sounds for Child-Robot Interaction. In *Social Robotics*, Miguel A. Salichs, Shuzhi Sam Ge, Emilia Ivanova Barakova, John-John Cabibihan, Alan R. Wagner, Álvaro Castro-González, and Hongsheng He (Eds.). Vol. 11876. Springer, Cham, Switzerland, 505–514. http://link.springer.com/10.1007/978-3-030-35888-4_47

[125] James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39, 6 (1980), 1161–1178. https://doi.org/10.1037/h0077714 Place: US Publisher: American Psychological Association.

[126] Hooman Aghaebrahimi Samani, Adrian David Cheok, and Owen Noel Newton Fernando. 2011. An affective interactive audio interface for Lovotics. *Computers in Entertainment* 9, 2 (July 2011), 1–14. https://doi.org/10.1145/1998376.1998377

[127] Hooman Aghaebrahimi Samani, Adrian David Cheok, Foo Wui Ngiap, Arjun Nagpal, and Mingde Qiu. 2010. Towards a formulation of love in human - robot interaction. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Viareggio, Italy, 94–99. https://doi.org/10.1109/ROMAN.2010.5598716

[128] Eleanor Sandry. 2015. Re-evaluating the Form and Communication of Social Robots: The Benefits of Collaborating with Machinelike Robots. *International Journal of Social Robotics* 7, 3 (June 2015), 335–346. https://doi.org/10.1007/s12369-014-0278-3

[129] Eleanor Sandry. 2017. Creative Collaborations with Machines. *Philosophy & Technology* 30, 3 (Sept. 2017), 305–319. https://doi.org/10.1007/s13347-016-0240-4

[130] Ravi Kiran Sarvadevabhatla, Victor Ng-Thow-Hing, and Sandra Okita. 2010. Extended duration human-robot interaction: Tools and analysis. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Viareggio, Italy, 7–14. https://doi.org/10.1109/ROMAN.2010.5598676

[131] Richard Savery, Amit Rogel, and Gil Weinberg. 2021. Emotion Musical Prosody for Robotic Groups and Entitativity. In *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Vancouver, BC, Canada, 440–446. https://doi.org/10.1109/RO-MAN50785.2021.9515314

[132] Richard Savery, Ryan Rose, and Gil Weinberg. 2019. Establishing Human-Robot Trust through Music-Driven Robotic Emotion Prosody and Gesture. In *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, New Delhi, India, 1–7. https://doi.org/10.1109/RO-MAN46459.2019.8956386

[133] Markus Schwenk and Kai O. Arras. 2014. R2-D2 Reloaded: A flexible sound synthesis system for sonic human-robot interaction design. In *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Edinburgh, UK, 161–167. https://doi.org/10.1109/ROMAN.2014.6926247

[134] Mojtaba Shahab, Alireza Taheri, Seyed Ramezan Hosseini, Mohammad Mokhtari, Ali Meghdari, Minoo Alemi, Hamidreza Pouretemad, Azadeh Shariati, and Ali Ghorbandaei Pour. 2017. Social Virtual Reality Robot (V2R): A Novel Concept for Education and Rehabilitation of Children with Autism. In *Proceedings of the RSI International Conference on Robotics and Mechatronics (ICRoM)*. IEEE, Tehran, Iran, 82–87. https://doi.org/10.1109/ICRoM.2017.8466148

[135] Mojtaba Shahab, Alireza Taheri, Mohammad Mokhtari, Azadeh Shariati, Rozita Heidari, Ali Meghdari, and Minoo Alemi. 2022. Utilizing social virtual reality robot (V2R) for music education to children with high-functioning autism. *Education and Information Technologies* 27, 1 (Jan. 2022), 819–843. https://doi.org/10.1007/s10639-020-10392-0

[136] Takanori Shibata, Kazuyoshi Inoue, and Robert Irie. 1996. Emotional robot for intelligent system-artificial emotional creature project. In *Proceedings of the IEEE International Workshop on Robot and Human Communication (RO-MAN)*. IEEE, Tsukuba, Japan, 466–471. https://doi.org/10.1109/ROMAN.1996.568881

[137] Takanori Shibata, Makoto Yoshida, and Junji Yamato. 1997. Artificial emotional creature for human-machine interaction. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC): Computational Cybernetics and Simulation*, Vol. 3. IEEE, Orlando, FL, USA, 2269–2274. https://doi.org/10.1109/ICSMC.1997.635205

[138] Koji Shibuya and Hiroyuki Ishimoto. 2018. Design Principles of Loudness to Express Bright and Dark Timbres for Violin-playing Robot. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Nanjing, China, 262–267. https://doi.org/10.1109/ROMAN.2018.8525660

[139] Masahiro Shiomi, Daisuke Sakamoto, Takayuki Kanda, Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita. 2011. Field Trial of a Networked Robot at a Train Station. *International Journal of Social Robotics* 3, 1 (Jan. 2011), 27–40. https://doi.org/10.1007/s12369-010-0077-4

[140] Moondeep C. Shrestha, Ayano Kobayashi, Tomoya Onishi, Hayato Yanagawa, Yuta Yokoyama, Erika Uno, Alexander Schmitz, Mitsuhiro Kamezaki, and Shigeki Sugano. 2016. Exploring the use of light and display indicators for communicating directional intent. In *Proceedings of the IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*. IEEE, Banff, AB, Canada, 1651–1656. https://doi.org/10.1109/AIM.2016.7577007

[141] Jorge Solis, Koichi Taniguchi, Takeshi Ninomiya, Klaus Petersen, Tetsuro Yamamoto, and Atsuo Takanishi. 2009. Implementation of an Auditory Feedback Control System on an Anthropomorphic Flutist Robot Inspired on the Performance of a Professional Flutist. *Advanced Robotics* 23, 14 (Jan. 2009), 1849–1871. https://doi.org/10.1163/016918609X12518783330207

[142] Florina Speth and Michael Wahl. 2014. Specifying Rhythmic Auditory Stimulation for Robot-assisted Hand Function Training in Stroke Therapy. In *Proceedings of the International Conference on Auditory Display (ICAD)*. The International Community for Auditory Display, New York, NY, USA, 8. https://doi.org/1853/52046

[143] Sophia C. Steinhaeusser, Philipp Schaper, and Birgit Lugrin. 2021. Comparing a Robotic Storyteller versus Audio Book with Integration of Sound Effects and Background Music. In *Companion of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Boulder, CO, USA, 328–333. https://doi.org/10.1145/3434074.3447186

[144] Rafał Stęgierski and Karol Kuczyński. 2014. The Perception of Humanoid Robot by Human. In *Image Processing and Communications Challenges*, Ryszard S. Choras (Ed.). Vol. 233. Springer, Heidelberg, Germany, 65–72. http://link.springer.com/10.1007/978-3-319-01622-1_8

[145] Alireza Taheri, Ali Meghdari, Minoo Alemi, Hamidreza Pouretemad, Pegah Poorgoldooz, and Maryam Roohbakhsh. 2016. Social Robots and Teaching Music to Autistic Children: Myth or Reality? In *Social Robotics*, Arvin Agah, John-John Cabibihan, Ayanna M. Howard, Miguel A. Salichs, and Hongsheng He (Eds.). Vol. 9979. Springer, Cham, Switzerland, 541–550. http://link.springer.com/10.1007/978-3-319-47437-3_53

[146] Marie Tahon and Laurence Devillers. 2016. Towards a Small Set of Robust Acoustic Features for Emotion Recognition: Challenges. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24, 1 (Jan. 2016), 16–28. https://doi.org/10.1109/TASLP.2015.2487051

[147] Adriana Tapus. 2009. Improving the Quality of Life of People with Dementia through the Use of Socially Assistive Robots. In *Proceedings of the Advanced Technologies for Enhanced Quality of Life (AT-EQUAL)*. IEEE, Iasi, Romania, 81–86. https://doi.org/10.1109/AT-EQUAL.2009.26

[148] Adriana Tapus, Cristian Tapus, and Maja Mataric. 2009. The role of physical embodiment of a therapist robot for individuals with cognitive impairments. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Toyama, Japan, 103–107. https://doi.org/10.1109/ROMAN.2009.5326211

[149] Keng Peng Tee, Rui Yan, Yuanwei Chua, and Zhiyong Huang. 2013. Audio-visual attention control of a pan-tilt telepresence robot. In *Proceedings of the International Conference on Control, Automation and Systems (ICCAS)*. IEEE, Gwangju, Korea (South), 827–832. https://doi.org/10.1109/ICCAS.2013.6704028

[150] Hamish Tennent, Dylan Moore, Malte Jung, and Wendy Ju. 2017. Good vibrations: How consequential sounds affect perception of robotic arms. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Lisbon, Spain, 928–935. https://doi.org/10.1109/ROMAN.2017.8172414

[151] Johann P Tissberger and György Wersényi. 2011. Sonification Solutions for Body Movements in Rehabilitation of Locomotor Disorders. In *Proceedings of the International Conference on Auditory Display (ICAD)*. The International Community for Auditory Display, Budapest, Hungary, 6. https://doi.org/1853/51751

[152] Vlad M. Trifa, Ansgar Koene, Jan Moren, and Gordon Cheng. 2007. Real-time acoustic source localization in noisy environments for human-robot multimodal interaction. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Jeju, South Korea, 393–398. https://doi.org/10.1109/ROMAN.2007.4415116

[153] Gabriele Trovato, Renato Paredes, Javier Balvin, Francisco Cuellar, Nicolai Baek Thomsen, Soren Bech, and Zheng-Hua Tan. 2018. The Sound or Silence: Investigating the Influence of Robot Noise on Proxemics. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Nanjing, China, 713–718. https://doi.org/10.1109/ROMAN.2018.8525795

[154] tyiannak. 2022. *GitHub - tyiannak/AUROS: A ROS framework for Audio Analysis*. National Centre of Scientific Research "Demokritos". Retrieved 2022-06-01 from https://hark.jp/

[155] René Tünnermann, Jan Hammerschmidt, and Thomas Hermann. 2013. Blended Sonification –Sonification For Casual Information Interaction. (July 2013). https://smartech.gatech.edu/handle/1853/51656

[156] Azumi Ueno, Kotaro Hayashi, and Ikuo Mizuuchi. 2019. Impression Change on Nonverbal Non-Humanoid Robot by Interaction with Humanoid Robot. In *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, New Delhi, India, 1–6. https://doi.org/10.1109/RO-MAN46459.2019.8956240

[157] Edwin van der Heide. 2010. Spatial Sounds (100dB at 100km/h) in the Context of Human Robot Personal Relationships. In *Proceedings of the International Conference on Human-Robot Personal Relationships (HRPR)*, Maarten H. Lamers and Fons J. Verbeek (Eds.), Vol. 59. Springer, Leiden, The Netherlands, 27–33. https://doi.org/10.1007/978-3-642-19385-9_4

[158] A. Vasilijevic, N. Miskovic, and Z. Vukic. 2013. Comparative assessment of human machine interfaces for ROV guidance with different levels of secondary visual workload. In *Proceedings of the Mediterranean Conference on Control and Automation (MED)*. IEEE, Platanias/Chania, Crete, Greece, 1292–1297. https://doi.org/10.1109/med.2013.6608886

[159] John Vilk and Naomi T. Fitter. 2020. Comedians in Cafes Getting Data: Evaluating Timing and Adaptivity in Real-World Robot Comedy Performance. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, Cambridge United Kingdom, 223–231. https://doi.org/10.1145/3319502.3374780

[160] Joachim von Zitzewitz, Patrick M. Boesch, Peter Wolf, and Robert Riener. 2013. Quantifying the Human Likeness of a Humanoid Robot. *International Journal of Social Robotics* 5, 2 (April 2013), 263–276. https://doi.org/10.1007/s12369-012-0177-4

[161] Klaus Weber, Hannes Ritschel, Ilhan Aslan, Florian Lingenfelser, and Elisabeth André. 2018. How to Shape the Humor of a Robot - Social Behavior Adaptation Based on Reinforcement Learning. In *Proceedings of the ACM International Conference on Multimodal Interaction (ICMI)*. ACM, Boulder, CO, USA, 154–162. https://doi.org/10.1145/3242969.3242976

[162] Klaus Weber, Hannes Ritschel, Florian Lingenfelser, and Elisabeth André. 2018. Real-Time Adaptation of a Robotic Joke Teller Based on Human Social Signals. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. International Foundation

for Autonomous Agents and Multiagent Systems, Richland, SC, USA, 2259–2261. https://doi.org/10.5555/3237383.3238141

[163] Wikimedia Foundation. 2022. *Earcon.* Wikimedia Foundation. Retrieved 2022-06-01 from https://en.wiktionary.org/wiki/earcon

[164] Wikimedia Foundation. 2022. *Jingle.* Wikimedia Foundation. Retrieved 2022-06-01 from https://en.wiktionary.org/wiki/jingle

[165] Wikimedia Foundation. 2022. *Paralanguage.* Wikimedia Foundation. Retrieved 2022-06-01 from https://en.wiktionary.org/wiki/paralanguage

[166] Wikimedia Foundation. 2022. *Sonification.* Wikimedia Foundation. Retrieved 2022-06-01 from https://en.wiktionary.org/wiki/sonification

[167] Wikimedia Foundation. 2022. *Vocable.* Wikimedia Foundation. Retrieved 2022-06-01 from https://en.wiktionary.org/wiki/vocable

[168] Wikimedia Foundation. 2022. *White noise.* Wikimedia Foundation. Retrieved 2022-09-30 from https://en.wiktionary.org/wiki/white_noise

[169] Hannah Wolfe, Marko Peljhan, and Yon Visell. 2020. Singing Robots: How Embodiment Affects Emotional Responses to Non-Linguistic Utterances. *IEEE Transactions on Affective Computing* 11, 2 (April 2020), 284–295. https://doi.org/10.1109/TAFFC.2017.2774815

[170] Guangyu Xia, Mao Kawai, Kei Matsuki, Mutian Fu, Sarah Cosentino, Gabriele Trovato, Roger Dannenberg, Salvatore Sessa, and Atsuo Takanishi. 2016. Expressive Humanoid Robot For Automatic Accompaniment. In *Proceedings of the Sound and Music Computing Conference (SMC).* Sound and Music Computing Network, Hamburg, Germany. https://doi.org/10.5281/zenodo.851327

[171] Seiji Yamada and Takanori Komatsu. 2006. Designing simple and effective expression of robot's primitive minds to a human. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* IEEE/RSJ, Beijing, China, 2614–2619. https://doi.org/10.1109/IROS.2006.281940

[172] Nobuhide Yamakawa, Toru Takahashi, Tetsuro Kitahara, Tetsuya Ogata, and Hiroshi G. Okuno. 2011. Environmental Sound Recognition for Robot Audition Using Matching-Pursuit. In *Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE) (Lecture Notes in Computer Science, Vol. 6704)*, Kishan G. Mehrotra, Chilukuri K. Mohan, Jae C. Oh, Pramod K. Varshney, and Moonis Ali (Eds.). Springer, Syracuse, NY, USA, 1–10. https://doi.org/10.1007/978-3-642-21827-9_1

[173] Michiya Yamamoto and Tomio Watanabe. 2007. Analysis by Synthesis of an Information Presentation Method of Embodied Agent Based on the Time Lag Effects of Utterance to Communicative Actions. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).* IEEE, Jeju, South Korea, 43–48. https://doi.org/10.1109/ROMAN.2007.4415051

[174] Jeong-Yean Yang and Dong-Soo Kwon. 2013. Feedback-based reasoning process for behavior selection during long-term interaction. In *Proceedings of the International Conference on Ubiquitous Robots and Ambient Intelligence (URAI).* IEEE, Jeju, Korea (South), 262–265. https://doi.org/10.1109/URAI.2013.6677363

[175] Selma Yilmazyildiz, Robin Read, Tony Belpeame, and Werner Verhelst. 2016. Review of Semantic-Free Utterances in Social Human–Robot Interaction. *International Journal of Human-Computer Interaction* 32, 1 (Jan. 2016), 63–85. https://doi.org/10.1080/10447318.2015.1093856

[176] Cristina Zaga, Roelof A.J. De Vries, Sem J. Spenkelink, Khiet P. Truong, and Vanessa Evers. 2016. Help-giving robot behaviors in child-robot games: Exploring Semantic Free Utterances. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI).* ACM/IEEE, Christchurch, New Zealand, 541–542. https://doi.org/10.1109/HRI.2016.7451846

[177] Lisa Zahray, Richard Savery, Liana Syrkett, and Gil Weinberg. 2020. Robot Gesture Sonification to Enhance Awareness of Robot Status and Enjoyment of Interaction. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).* IEEE, Naples, Italy, 978–985. https://doi.org/10.1109/RO-MAN47096.2020.9223452

[178] Brian J. Zhang, Knut Peterson, Christopher A. Sanchez, and Naomi T. Fitter. 2021. Exploring Consequential Robot Sound: Should We Make Robots Quiet and Kawaii-et?. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* IEEE/RSJ, Prague, Czech Republic, 3056–3062. https://doi.org/10.1109/IROS51168.2021.9636365

[179] Brian J. Zhang, Noel Sigafoos, Rabecka Moffit, Lili Adams, Ibrahim Syed, Jason Fick, and Naomi T. Fitter. 2022. SonifyIt: Towards Transformative Sound for All Robots. *accepted for publication in the IEEE Robotics and Automation Letters (RA-L)* (2022).

[180] Brian J. Zhang, Nick Stargu, Samuel Brimhall, Lilian Chan, Jason Fick, and Naomi T. Fitter. 2021. Bringing WALL-E out of the Silver Screen: Understanding How Transformative Robot Sound Affects Human Perception. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA).* IEEE, Xi'an, China, 3801–3807. https://doi.org/10.1109/ICRA48506.2021.9562082

[181] Ruimin Zhang, Jaclyn Barnes, Joseph Ryan, Myounghoon Jeon, Chung Hyuk Park, and Ayanna M. Howard. 2016. Musical Robots For Children With ASD Using A Client-Server Architecture. In *Proceedings of the International Conference on Auditory Display (ICAD).* The International Community for Auditory Display, Canberra, Australia, 83–89. https://doi.org/10.21785/icad2016.007

A REVIEWED PAPERS BY FUNCTION

Function	Year	Authors
Sound source recognition	2011	N. Yamakawa et al. [172]
	2012	M. Janvier et al. [61]
	2012	T. Araki et al. [2]
	2014	M. Janvier et al. [62]
	2016	H. Kudo et al. [74]
	2016	T. Giannakopoulos and G. Siantikos [38]
	2018	L. Grama and C. Rusu [45]
	2019	L. Martínez-Villaseñor and H. Ponce [86]
	2021	L. Muscar et al. [99]
	2022	M. A. Maheux et al. [83]
Paralanguage recognition	2007	M. Yamamoto and T. Watanabe [173]
	2010	H. A. Samani et al. [127]
	2010	Y. Lin et al. [81]
	2011	H. A. Samani et al. [126]
	2011	H. Knight [69]
	2011	M. Shiomi et al. [139]
	2012	J. S. Park et al. [111]
	2013	D. K. Limbu et al. [80]
	2016	M. Tahon and L. Devillers [146]
	2020	J. Vilk and N. T. Fitter [159]
	2020	S. Jaiswal et al. [60]
Music recognition	2007	M. P. Michalowski et al. [95]
	2009	A. Tapus [147]
	2009	A. Tapus et al. [148]
	2009	M. P. Michalowski et al. [96]
	2012	A. Lim et al. [79]
	2012	G. Hoffman [54]
	2012	J. L. Oliveira et al. [110]
	2012	J. S. Park et al. [111]
	2013	G. Hoffman and K. Vanunu [56]
	2015	G. Ince et al. [57]
	2015	H. Peng et al. [113]
	2016	A. F. Azmin et al. [3]
	2016	A. Taheri et al. [145]
	2016	G. Hoffman et al. [55]
	2016	G. Xia et al. [170]
	2017	E. Sandry [129]

Function	Year	Authors
Music recognition	2017	M. Shahab et al. [134]
	2020	S. Chakraborty and J. Timoney [23]
	2020	W. K. N. Hansika et al. [48]
	2021	F. Ciardo et al. [25]
	2021	G. Ince et al. [58]
	2021	J. A. Barnes et al. [4]
	2021	M. Krzyżaniak [73]
	2022	M. Shahab et al. [135]
Sound source localization & separation	1996	T. Shibata et al. [136]
	1997	T. Shibata et al. [137]
	1999	R. A. Brooks et al. [15]
	2002	H. G. Okuno et al. [107]
	2003	H. G. Okuno and K. Nakadai [109]
	2003	H. G. Okuno et al. [108]
	2003	H. G. Okuno et al. [105]
	2003	H. G. Okuno et al. [104]
	2005	L. Błażejewski [17]
	2005	M. Bennewitz et al. [7]
	2007	V. M. Trifa et al. [152]
	2009	H. D. Kim et al. [68]
	2010	R. K. Sarvadevabhatla et al. [130]
	2011	H. G. Okuno et al. [106]
	2012	N. Masuyama et al. [87]
	2013	K. P. Tee et al. [149]
	2014	E. Martinson and V. Yalla [85]
	2014	R. Stęgierski and K. Kuczyński [144]
	2014	S. E. Fotinea et al. [33]
	2018	L. Grama and C. Rusu [45]
	2018	W. He et al. [51]
	2021	J. Fan et al. [28]
	2022	M. A. Maheux et al. [83]
	2022	U. Maniscalco et al. [84]
Other sound perception	2006	D. Brock and E. Martinson [13]
	2008	N. A. Mirza et al. [97]
	2010	C. Kroos et al. [72]
	2018	A. Nijholt [101]
	2018	K. Weber et al. [161]
	2018	K. Weber et al. [162]

Function	Year	Authors
Other sound perception	2022	U. Maniscalco et al. [84]
Consequential robot sound	2013	J. von Zitzewitz et al. [160]
	2017	D. Moore et al. [98]
	2017	H. Tennent et al. [150]
	2018	E. Frid et al. [36]
	2020	L. Boos and L. Moshkina [11]
	2020	T. Izui and G. Venture [59]
	2021	B. J. Zhang et al. [178]
Transformative robot sound	2010	E. van der Heide [157]
	2014	M. Joosse et al. [67]
	2017	J. Bellona et al. [6]
	2017	L. Dahl et al. [26]
	2018	E. Cha et al. [20]
	2018	G. Trovato et al. [153]
	2021	B. J. Zhang et al. [180]
	2021	J. S. Lee et al. [78]
	2022	E. Frid and R. Bresin [35]
Emotional robot sound	2006	S. Yamada and T. Komatsu [171]
	2010	R. Read and T. Belpaeme [120]
	2011	H. A. Samani et al. [126]
	2012	R. Read and T. Belpaeme [116]
	2013	D. K. Limbu et al. [80]
	2013	J. Y. Yang and D. Kwon [174]
	2013	R. Read and T. Belpaeme [117]
	2013	S. Pourmehr et al. [115]
	2014	M. Schwenk and K. O. Arras [133]
	2014	R. Read and T. Belpaeme [118]
	2014	R. Read and T. Belpaeme [119]
	2015	E. Sandry [128]
	2015	L. Boccanfuso et al. [8]
	2016	A. F. Azmin et al. [3]
	2016	H. Hastie et al. [50]
	2016	R. Zhang et al. [181]
	2016	S. Yilmazyildiz et al. [175]
	2017	E. Jeong et al. [64]
	2017	J. Fernandez De Gorostiza luengo et al. [29]
	2018	D. Löffler et al. [82]
	2018	E. Frid et al. [36]

Function	Year	Authors
Emotional robot sound	2018	K. Weber et al. [161]
	2018	K. Weber et al. [162]
	2019	M. R. Frederiksen and K. Stoey [34]
	2019	R. Savery et al. [132]
	2019	S. Rossi et al. [124]
	2020	A. B. Latupeirissa et al. [77]
	2020	H. R. M. Pelikan et al. [112]
	2020	H. Wolfe et al. [169]
	2021	B. J. Zhang et al. [180]
	2021	J. S. Lee et al. [78]
	2021	R. Savery et al. [131]
	2021	S. C. Steinhaeusseret al. [143]
Functional robot sound	1999	A. Camurri et al. [18]
	2002	G. Johannsen [66]
	2003	T. Hermann et al. [53]
	2007	E. C. Haas [46]
	2007	K. Kobayashi et al. [71]
	2010	E. van der Heide [157]
	2010	R. Read and T. Belpaeme [120]
	2011	A. Mertens et al. [94]
	2011	J. P. Tissberger and G. Wersenyi [151]
	2013	A. Vasilijevic et al. [158]
	2013	K. L. Koay et al. [70]
	2014	F. Speth and M. Wahl [142]
	2014	K. Fischer et al. [30]
	2014	K. Fischer et al. [31]
	2014	M. Schwenk and K. O. Arras [133]
	2014	S. Bökesoy [16]
	2016	C. Zaga et al. [176]
	2016	E. Cha and M. Matarić [22]
	2016	E. Cha et al. []
	2016	E. Florentine et al. [21]
	2016	M. C. Shrestha et al. [140]
	2016	S. Lakhmani et al. [75]
	2016	S. Yilmazyildiz et al. [175]
	2017	E. Jeong et al. [64]
	2018	G. Bolano et al. [9]
	2019	A. Ueno et al. [156]
	2019	H. Ritschel et al. [121]

Function	Year	Authors
Functional robot sound	2019	P. Jin et al. [65]
	2020	J. Okimoto and N. Niitsuma [103]
	2020	L. Boos and L. Moshkina [11]
	2021	J. S. Lee et al. [78]
	2022	U. Maniscalco et al. [84]
Music synthesis	2009	B. Robins et al. [122]
	2009	J. Solis et al. [141]
	2010	R. Nikolaidis and G. Weinberg [102]
	2012	A. Lim et al. [79]
	2014	L. McCallum and P. W. McOwan [88]
	2015	G. Ince et al. [58]
	2015	L. McCallum and P. W. McOwan [89]
	2016	A. Taheri et al. [145]
	2016	G. Xia et al. [170]
	2017	E. Sandry [129]
	2017	M. Shahab et al. [134]
	2018	K. Shibuya and H. Ishimoto [138]
	2018	L. McCallum and P. W. McOwan [90]
	2020	S. Chakraborty and J. Timoney [23]
	2021	F. Ciardo et al. [25]
	2021	G. Ince et al. [58]
	2021	M. Krzyżaniak [73]
	2021	T. R. P. Pessanha et al. [114]
	2022	M. Shahab et al. [135]
Other sound creation	2006	J. F. Gorostiza et al. [44]
	2012	K. S. Chun et al. [24]
	2017	R. Agrigoroaie and A. Tapus [1]

Table 4. Articles included in the review, arranged by the taxonomy of form proposed in Section 3.1, year of publication, and alphabetical order of the first author surname.