Perceptual Evaluation of Educational Robots' Consequential Sounds

Ivana Storjak[1] and Ana Sovic Krzic[1]

[1] University of Zagreb Faculty of Electrical Engineering and Computing, Zagreb, Croatia

ivana.storjak@fer.hr,
ana.sovic.krzic@fer.hr

Abstract. Educational robots are no longer solely used in extracellular activities but also in formal education. To reach learning goals, multiple factors should be observed when choosing such a tool to augment learning. This study reasons that the sounds robots produce intentionally and consequentially bring up the noise level in the classroom that might not benefit learning, and the quality of such sounds can impact the acceptability of the robot. Therefore, the consequential sounds of six different educational robots (mBot, LEGO Mindstorms EV3, Cubelets, Thymio, Codey Rocky, and Jimu) were studied on affective sound quality (commonness of sound, pleasantness, likeness, and annoyance), and perception related to specific quality attributes of the robot producing the sound (robot's vitality, quality, robustness, precision, and safety). The treatment was organized in a form of a live listening test, and sound impressions were collected by using nine five-point bipolar Likert items. Additionally, pupils were asked to rank the robots based on their respective sounds and appearance separately. As a result, observable differences were noted. mBot was rated most positively, followed by Thymio, and the worst ratings were attributed to the Jimu robot, probably due to the interaction of its caterpillar treads with the wooden surface. On the other hand, Jimu had the best ranking on behalf of its appearance. Cumulatively, based solely on the sounds, most pupils think that robots are safe, somewhat precise, robust, and working properly. However, the sounds are perceived as somewhat bothering, unpleasant and unusual and could distract pupils while learning.

Keywords: Consequential sounds · Educational robots · Learning environment · Listening test · Perceptual evaluation · Robot design

1 Introduction

For effective technology integration in educational settings, understanding learning theories and technology integration models is crucial. Learning theories imply different roles of technology in learning, while the technology integration models determine the process and objectives of the integration [1]. Regardless of differences imposed by the adopted approach, the purpose of technology implementation is to augment the learning process and reach learning goals.

In recent years, the utilization of various technological tools such as robots in learning environments has been explored from multiple disciplines and research directions. From a pedagogical aspect, the novelty and interactiveness of robots were linked with higher motivation and engagement, which positively affect students' performance and achievements. Still, to completely understand the underlying factors that influence human-robot interaction (HRI), the psychological dimension should also be considered.

One of such factors that affect acceptance is the perception of robots. So far, research regarding the perception of robots was mainly concerned with robot design features like physical attributes [2] and personality features [3] but not so much with the acoustic aspects of robots [4].

Auditory-enhanced interfaces entail sound design and sonic interaction design (SID), which complement the product experience. Still, not all robot-related sound events are intentional, hence, designed to convey a message, provide feedback, alarm the user [5], and are produced using a loudspeaker or a piezo element [6]. Another type of sounds are consequential sounds, which arise from operating a robot (sound radiated from the robot's mechanics or interaction with the surface) [6] and from the construction of a robot. These can also be accompanied by vibration.

Therefore, the implementation of robots as learning tools implies the addition of noise in a shared learning environment, e.g., intentional, and consequential sounds of robots and potentially noise coming from students' engagement in the robot activities, e.g., babble. As sounds impact performance in the classroom [7], suitability of robots for a designated context should be examined from an auditory perspective.

Listening tests are broadly used to evaluate the sound quality of various devices. [8] describes theory, methodology, and application of perceptual sound evaluation in detail. In [9], sound quality of hand-held tools was investigated on four groups of subjective parameters related to describing the sound itself, the perception of sound and its effect on humans, properties of the device, and a desire to buy a product. 11-point category scale was chosen to evaluate pleasantness, safety, robustness, quality, proper functioning, power, and value.

Similarly, [4] studied the perception of robot servo motors sounds due to the influence of robots on the perception but without giving a context of HRI. The authors chose seven subjective parameters that are vital for sound's acceptability, such as precision, preference, roughness, strength, pleasantness, annoyance, and expensiveness.

Another study of interest is [10], which deals with the effect of a robot's noise on proxemics and how it can be eliminated by evaluating the personal perception of sound and safety in the environment. Here, 7-points semantic differential scales (calming-agitating, gentle-harsh, soft-hard, smooth-rough, friendly-unfriendly, pleasant-unpleasant, regular-irregular), 5-point Likert scales, and eight emotion-related attributes were utilized (worried, fearful, anxious, angry, sad, disappointed, guilty, and disgusted).

In [11], stimuli consisted of Pepper's robot servomotor noises, sawtooth or other complex designed sounds and was presented in a form of audio, audio with video, or video. Ratings on four semantic opposites related to four quadrants of the circumplex space of emotions, and five related to the quality of robot behaviour such as pleasantness, typicality, efficiency, likability, and trust. Although [9] notes that the selection of subjective parameters is product-specific, it seems to be also related to the study's objective.

In our study, we compare six educational robots based on their sounds using subjective opinions of the targeted audience (pupils): affective measures (commonness of sound, pleasantness, likeness, and annoyance), as well as evaluation of robots' quality attributes based on its sound (robot's vitality, quality, robustness, precision, and safety).

2 Methodology

2.1 Participants

Data for this study were acquired at the summer robotics camp "Petica 2021" in July 2021. For learning activities, pupils were divided into three classrooms based on age and previous experience with programming and robotics. To practice programming beginners' class utilized LEGO Mindstorms EV3 kit, the experienced class used LEGO SPIKE Prime, and in the advanced class, the VIDI X (ESP32 based microcontroller) was programmed. The treatment took place after four days of camp. Parents' active consent was collected before the beginning of the camp, and all the camp participants voluntarily participated in the study. The sample included 17 pupils (10-15 years old, F 5 M 12). Pupils were a target group of this study due to being users of this technology in educational settings.

2.2 Materials

Six different robots were included in this exploratory study. Some of the used robots require assembly, such as Makeblock mBot STEM Education Coding Robot Kit pink version[1], LEGO Mindstorms EV3[2], UBTECH Jimu Cosmos kit[3], and Cubelets Boundless Builder Pack[4] while Thymio[5] and Makeblock Codey Rocky[6] are ready-to-use straight out of the box. The goal appearance was a vehicle (Fig. 1b). The Bulldozer design was assembled with LEGO Mindstorms EV3 kit, the Rover model with Jimu kit and Cubelets were assembled using only six blocks (four motors, battery, and ultrasonic sensor).

---

1 https://www.makeblock.com/steam-kits/mbot
2 https://www.lego.com/en-us/product/lego-mindstorms-ev3-31313
3 https://www.ubtrobot.com/collections/jimu-robot-kits
4 https://modrobotics.com/
5 https://www.thymio.org/
6 https://www.makeblock.com/steam-kits/codey-rocky

For the treatment, robots were programmed to perform a simple forward-backward movement in a loop. Speed was set to the middle value of the span, and the duration of the drive in each direction was 2 seconds. Although this study is concerned with the consequential sounds of educational robots, apart from a consequential component of a sound, Codey Rocky's sound also had an intentional component, an occasional high-pitched sound coming from the speaker.

2.3 Procedure

Investigation of pupils' subjective impressions of educational robots' sounds was performed by administering a survey and conducting a live listening test. Items in a survey with nine bipolar Likert items were in a form of a statement and concerned with four affective measurements of the sound quality and five scales with specific anchoring attributes related to the quality of a robot producing the sound. The statement regarding the affective quality of sound was: "The sound I have heard (is) …" and for the quality attributes of a robot: "According to the sound I have heard, I think the robot is … ". Where possible, concerning language restrictions, anchoring attributes were exact opposite words. Affective measures comprehended commonness of sound (usual – unusual), pleasantness (pleasant – unpleasant), likeness (like – dislike), and annoyance (does bother – does not bother). On the other hand, evaluation of a robot's quality attributes based on its sound comprised robot's vitality (working properly – broken), quality (of quality – not of quality), robustness (robust – fragile), precision (precise – not precise) and safety (safe – dangerous). The balance between two opposed extremes, first and second, was described symmetrically with a continuum of categorical options using the following wording: completely first, somewhat first, neither first nor second (neutral), somewhat second, and completely second, hence unbiased end-to-end options were utilized.

The procedure of conducting a listening test comprised familiarizing subjects with the study and asking for participation. After giving consent, pupils were seated side by side in a line to be at a proximate distance to a robot (see Fig. 1a). To exclude the influence of a visual stimuli bias when rating sounds (auditory stimuli), hence the influence of the robot's appearance on the ratings, pupils were facing the opposite direction. Also, to exclude the interference of robot sound with background noise such as babble and provide stable listening conditions, pupils were asked to be silent. Each stimulus in this live listening test was presented not as a recording but by starting each of the included robots one at a time. The first sound presented belonged to mBot, followed by EV3, Jimu, Codey Rocky, Thymio, and lastly, Cubelets. Regarding fidelity, the rooms where treatment took place were somewhat representative of a classroom listening space, and a wooden surface was used (floor or desk), which is a common use case scenario for utilizing robots in a conventional classroom.

After rating robots on the nine Likert items, pupils were asked to rank the robots based on the sound from most liked to least liked. With this step, the listening test was over, and pupils were asked to turn around to rank order robots based on their appearance. For both ranking tasks, stimuli, hence sound and appearance, were presented one by one in the same order as for the rating task. The described procedure was repeated separately for each of the three classes of pupils to maintain better listening conditions.

The aspects of interest of this study are summarized with the following research questions:

RQ1: What are the similarities and differences between impressions of robots' sounds?

RQ2: Which robot is superior regarding affective measures of the sound quality?

RQ3: Which robot is superior regarding robots' quality measures based on its sound?

RQ4: What are the similarities and differences between rankings of the sound and appearance of the robots?

RQ5: From a sound rating perspective, are robots acceptable for the learning environment (positive connotation of overall ratings)?

To answer these questions, data were observed using descriptive statistics measures such as frequency distribution, measures of central tendency, and variability. Frequency distribution describes the occurrence of different values in a data set. Applied to this study, it summarizes how many pupils rated the sound of a specific robot with a specific option on a respective Likert item. Further, considering the ordinal nature of data, measures of central tendency observed were mode and median. Mode represents the most frequent value among the observed values, which, in terms of this study, represents the most frequent option used regarding a specific statement (and robot). Further, the median denotes the middle number in an ordered list of values and is not affected by an asymmetrical distribution that affects mode. And lastly, variability was observed. For a given dataset, if any of the rating options is not present in the data, pupils' opinions are more in accord within the group.

Fig. 1. a) Pupils participating in the listening test, **b)** Post treatment interaction with robots.

3 Results and Discussion

Firstly, the sounds of the robots were evaluated on a scale going from *completely unusual* to *completely usual* (Fig. 2). According to pupils' ratings, the most unusual sound was the sound of Jimu, 82% of ratings were *completely unusual* or *somewhat unusual*. On the other hand, the most usual sound was the sound of mBot. Interestingly, Thymio had an almost uniform distribution of the ratings from *somewhat unusual* to *completely usual*.

On an item questioning pleasantness of sound (Fig. 3), Thymio had a mode in the *completely pleasant* category, and it was not at all rated as *completely unpleasant*. The slightly worse results were obtained for the mBot. On the opposite side, Jimu and EV3 did not get any *completely pleasant* ratings.

Most pupils rated the sound of Thymio as the sound they liked the most (Fig. 4). At the same time, they had divided opinions about which sound they disliked the most: Cubelets, Codey Rocky, Jimu, or EV3, where EV3 had mode in the *completely dislike* category. However, all of them had a median in the neutral category.

mBot was the robot whose sounds bothered pupils the least (with a mode in the *completely does not bother* category). Also, Thymio was rated similarly but with more neutral ratings (Fig. 5). Sounds that bothered pupils the most belonged to the EV3 and Jimu, with 76% in the *completely does bother* or *somewhat does bother* categories.

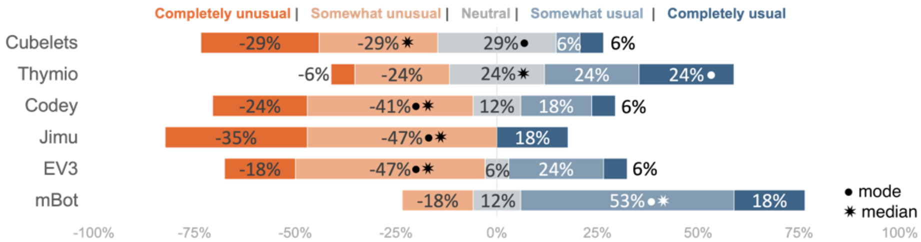

Fig. 2. Commonness of robots' sounds.

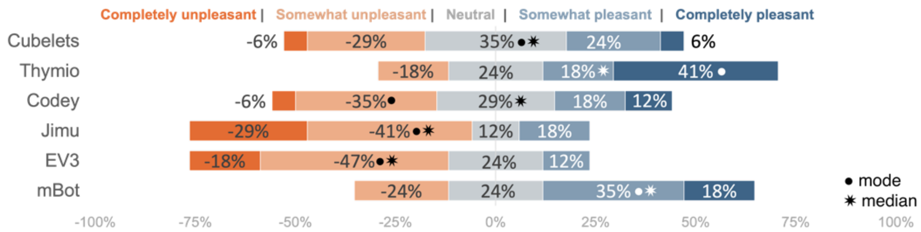

Fig. 3. Pleasantness of robots' sounds.

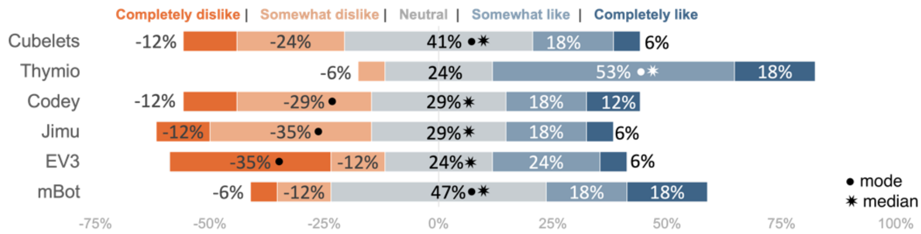

Fig. 4. Likeness of robots' sounds.

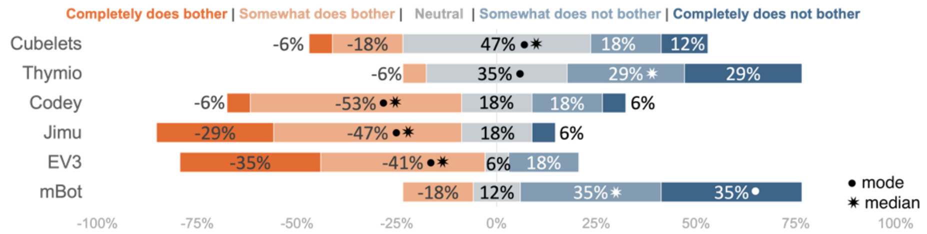

Fig. 5. Annoyance of robots' sounds.

The following results (Fig. 6 - Fig. 10) are concerned with the perception of robots' quality attributes based on the respective sounds. Fig. 6 gives results for an impression of robots' vitality based on the sound. Thymio and mBot showed very similar results, with a mode in the *completely working properly* category and a median in the *somewhat working properly* category. On the opposite, Jimu and Cubelets had mode in the *somewhat broken* category.

Thymio showed total superiority in the perceived quality of a robot based on the sound (Fig. 7), with 71% of ratings in the *somewhat of quality* category. Cubelets and Codey Rocky had ratings spread in all categories, but the mode was *somewhat not of quality*, whereas Cubelets had more of the *completely not of quality* ratings.

Most of the robots (Thymio, Jimu, EV3, mBot) were equally perceived as robust (Fig. 8), and EV3 even had a mode in the *completely robust* category. On the contrary, Cubelets and Codey Rocky had more *completely fragile* ratings than other robots, with a median in the neutral category.

For the precision of a robot based on the sound (Fig. 9), mBot showed supremacy with no negative ratings (*completely* or *somewhat not precise*). Thymio also showed superior results, with 88% of the ratings in categories *somewhat* precise and *completely precise*. Most negative ratings pertained to Jimu.

Similar results were obtained regarding the perceived safety of a robot based on the sound (Fig. 10), where Thymio had mode and median in the *completely safe* category and 88% of ratings in the *somewhat safe* and *completely safe* categories. At the same time, mBot had 82% in those categories. Compared to other robots, Jimu sounded most dangerous with a mode and median in the *neutral* category.

If we take into consideration all given analyses (RQ1), Jimu produced a sound that is most unusual and unpleasant and bothered pupils the most (RQ2). Based on the sound, pupils thought that the robot could be broken, not precise, and the most dangerous (RQ3). However, Jimu was perceived as robust, similarly to Thymio, mBot, and EV3.

On the opposite, Thymio and mBot possess the most usual and pleasant sounds that do not bother pupils (RQ2). Based on the sound, the robots work properly, are of quality (Thymio), precise and safe (RQ3). Even more, the sound of the Thymio robot was liked the most.

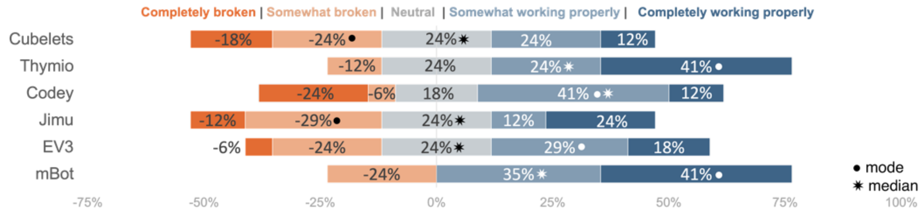

Fig. 6. Robots' vitality based on the sound.

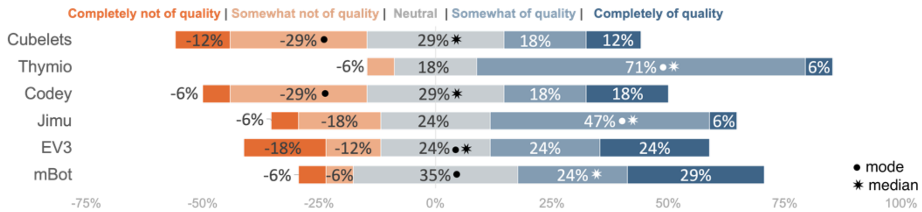

Fig. 7. Robots' quality based on the sound.

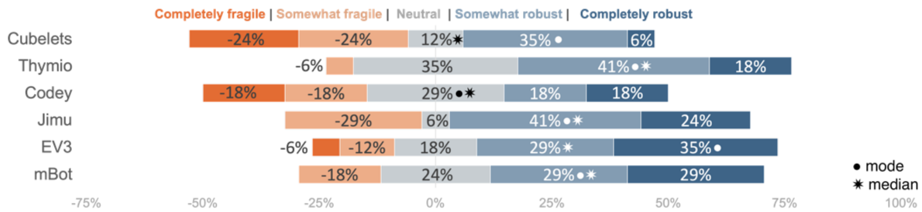

Fig. 8. Robots' robustness based on the sound.

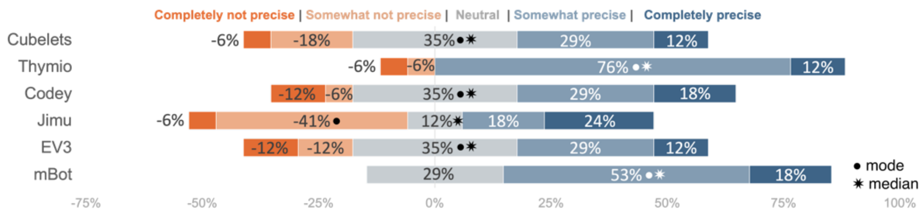

Fig. 9. Robots' precision based on the sound.

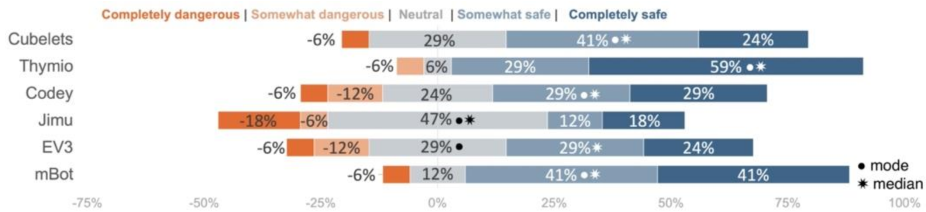

Fig. 10. Robots' safety based on the sound.

The following results are related to the ranked preference of the robots based on sound and appearance (RQ4). The analyses of ranking order data were performed on a sample of 15 pupils. Two male pupils were excluded due to missing data.

Ranked preferences of robots based on sound (Fig. 11) show that the best sound belongs to mBot with a total of 80% first and second choices. Thymio produces the second-best consequential sound with the same number of first choices as mBot, but with fewer second choices. Cubelets are the third choice, Codey Rocky the fourth, EV3 fifth, and Jimu is the last choice. The results of the ranked preferences coincide with the results of pupils' perceptual evaluations of the sounds produced by robots (Fig. 2 - Fig. 10).

Finally, after the sound treatment, pupils saw robots and ranked them by appearance (Fig. 12). Interestingly, the appearance rankings are in total opposition to the sound rankings. The most first choices were attributed to Jimu, then mBot, and Cubelets. EV3 had the most second choices, and in total is ranked in fourth place. The fifth is Codey Rocky, while the Thymio is the last choice. A possible explanation for Codey Rocky and Thymio is simplicity and a neutral, white-colored design ready-to-use straight out of the box. Oppositely, Jimu had caterpillars, a more colorful and complex appearance, yet it is possible that the sound of caterpillars on the wooden desk pushed Jimu to the last place in ranking based on the sound. The only robot that was in the top choices in both analyses is mBot. Although this study utilized a pink version instead of the regular blue that is commonly found in school, the sample of pupils consisting predominantly of boys was not bothered by the color.

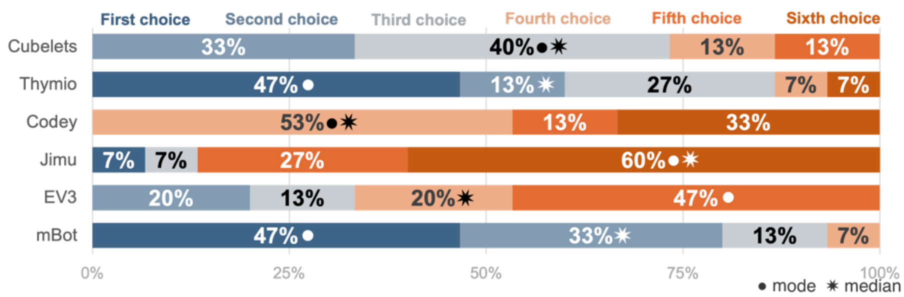

Fig. 11. Ranked preferences of robots based on sound.

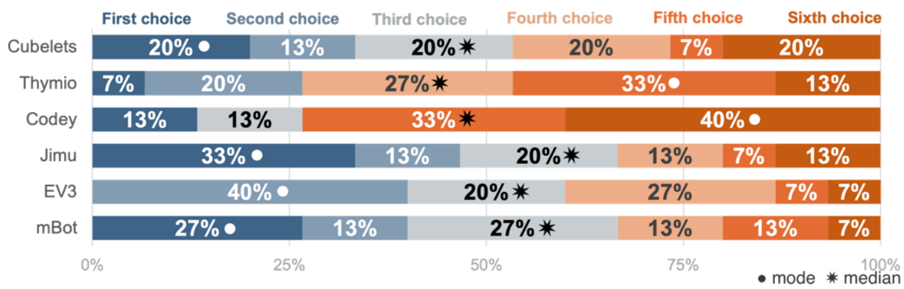

Fig. 12. Ranked preferences of robots based on appearance.

Overall (Fig. 13), most pupils think that robots are safe (mode in the *completely safe* and median in the *somewhat safe* category), somewhat precise, robust, and working properly. Safety-related results might be related to pupils knowing that they were hearing educational robots (and not industrial or military robots). Since pupils had never seen or heard most of the robots before, the mode and median of their sounds were *somewhat unusual*. At the same time, some of the robots produce unpleasant sounds that bother pupils and could distract them during the learning process (modes are in the *somewhat unpleasant* and *somewhat bothering* and median in the neutral category). Since they liked half of the robots and disliked the rest of them, on the scale between dislike and like, mode and median are in the *neutral* category. Median is in the *neutral* category also for the quality of the robots based on the sound. Although those are cumulative results for the six different educational robots, all distributions related to the quality of a robot producing the sound are tilted towards positive perceptions, which leads to the conclusion that the sounds of robots are somewhat well adapted to the target audience. However, the sounds somewhat bother pupils, are unpleasant and unusual, and could distract pupils from learning.

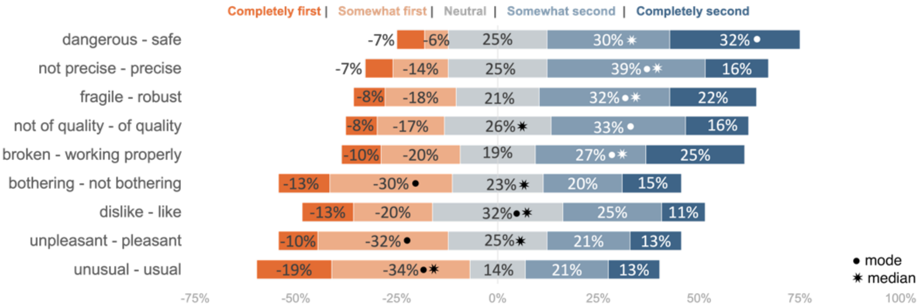

Fig. 13. Cumulative ratings of all tested robots based on the sound.

4 Conclusion and future work

The results of this study are relevant to robotics, educators, the sound engineering community and acousticians, and researchers in the mentioned fields with a mutual goal of understanding and contributing to the improvement of HRI.

Similarities and differences between impressions of robots' sounds (RQ1) were analyzed based on their affective measures (commonness of sound, pleasantness, likeness, and annoyance) and perception of robots' quality attributes based on their sound (robot's vitality, quality, robustness, precision, and safety). mBot and Thymio showed superiority regarding affective measures of the sound quality (RQ2) and regarding robots' quality measures based on their sound (RQ3). Similarly, pupils ranked mBot as a first choice and Thymio as a second choice in ranking robots by their sound. Contrary, Jimu was ranked first by robot appearance and Thymio as the last one due to its appearance simplicity (RQ4). From a sound rating perspective, the sounds of robots lead to a conclusion that educational robots are somewhat safe, precise, robust, of quality, and working properly. However, the sounds are somewhat unpleasant, unusual, and bother pupils, therefore could distract them from learning (RQ5). Even though consequential sounds are not so easy to alter as intentional sounds, both types bring the noise to the learning environment.

Although the presented results highlight differences in perception of robots' consequential sounds, there are several potential limitations of this study that could be addressed in future work to reach the generality of results. First concerns the sampling method, hereby used convenience sample could be substituted with a representative sample of pupils at a certain educational level or age span. This approach would increase the sample size and statistical significance of the results, therefore accurately reflecting the perception of the studied sounds within the population of interest. Also, multiple countries could be included using recordings. Another improvement for future work would be conducting the research through listening tests with a pairwise comparison. With a greater dataset, a predictive model could be built by combining acoustical and psychoacoustical analyses and the users' feedback. Using the Likert items in this study confirmed that this type of survey is a time-effective instrument to get feedback after an event. Also, the ease of delivering feedback was reflected with a 100% completion rate with no missing data. Still, although fewer options in the scale make it more user-friendly, this approach captures less detail. Therefore, future work might benefit from using 7-point Likert items.

Despite the listed limitations, the results presented in this study suggest practical implications of robot sound awareness in the context of a classroom. To conclude, robots with rubber tires, subtle in volume and with moderate buzzing or cracking, and smoother gradual momentum during starting or stopping the motors were preferred.

Acknowledgment

This work has been supported by Croatian Science Foundation under the project UIP-2017-05-5917 HRZZ-TRES. The work of doctoral student Ivana Storjak has been fully supported by the "Young researchers' career development project – training of doctoral students" of the Croatian Science Foundation ESF-DOK-2018-01.

References

[1] A. Ottenbreit-Leftwich and R. Kimmons, 'The K-12 Educational Technology Handbook', p. 415.

[2] T. Komatsu and S. Yamada, 'How Does the Agents' Appearance Affect Users' Interpretation of the Agents' Attitudes: Experimental Investigation on Expressing the Same Artificial Sounds From Agents With Different Appearances', *International Journal of Human-Computer Interaction*, vol. 27, no. 3, pp. 260–279, Feb. 2011, doi: 10.1080/10447318.2011.537209.

[3] S. Woods, 'Exploring the design space of robots: Children's perspectives', *Interacting with Computers*, vol. 18, no. 6, pp. 1390–1418, Dec. 2006, doi: 10.1016/j.intcom.2006.05.001.

[4] D. Moore, H. Tennent, N. Martelaro, and W. Ju, ‘Making Noise Intentional: A Study of Servo Sound Perception’, in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, Vienna Austria, Mar. 2017, pp. 12–21. doi: 10.1145/2909824.3020238.

[5] E. Jeong, G. H. Kwon, and J. So, ‘Exploring the taxonomic and associative link between emotion and function for robot sound design’, p. 3.

[6] L. Langeveld, R. van, R. Jansen, and E. Ozc, ‘Product Sound Design: Intentional and Consequential Sounds’, in *Advances in Industrial Design Engineering*, D. Coelho, Ed. InTech, 2013. doi: 10.5772/55274.

[7] J. E. Dockrell and B. M. Shield, ‘Acoustical barriers in classrooms: the impact of noise on performance in the classroom’, *British Educational Research Journal*, vol. 32, no. 3, pp. 509–525, Jun. 2006, doi: 10.1080/01411920600635494.

[8] S. Bech and N. Zacharov, *Perceptual Audio Evaluation-Theory, Method and Application: Bech/Perceptual Audio Evaluation-Theory, Method and Application*. Chichester, UK: John Wiley & Sons, Ltd, 2006. doi: 10.1002/9780470869253.

[9] M. Horvat, H. Domitrovi, and K. Jambrošić, ‘Sound Quality Evaluation of Hand-Held Power Tools’, *Acta Acustica united with Acustica*, vol. 98, no. 3, pp. 487–504, May 2012, doi: 10.3813/AAA.918532.

[10] G. Trovato *et al.*, ‘The Sound or Silence: Investigating the Influence of Robot Noise on Proxemics’, in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Nanjing, Aug. 2018, pp. 713–718. doi: 10.1109/ROMAN.2018.8525795.

[11] A. B. Latupeirissa, C. Panariello, and R. Bresin, ‘Exploring emotion perception in sonic HRI’, 2020, pp. 434–441.