International Journal of Social Robotics
https://doi.org/10.1007/s12369-022-00933-7

Design, Manufacture, and Acceptance Evaluation of APO: A Lip-syncing Social Robot Developed for Lip-reading Training Programs

Alireza Esfandbod[1] · Ahmad Nourbala[1] · Zeynab Rokhi[1] · Ali F. Meghdari[1,2] · Alireza Taheri[1] · Minoo Alemi[1,3]

Accepted: 27 September 2022

Abstract

Lack of educational facilities for the burgeoning world population, financial barriers, and the growing tendency in favor of inclusive education have all helped channel a general inclination toward using various educational assistive technologies, e.g., socially assistive robots. Employing social robots in diverse educational scenarios could enhance learners' achievements by motivating them and sustaining their level of engagement. This study is devoted to manufacturing and investigating the acceptance of a novel social robot named APO, designed to improve hearing-impaired individuals' lip-reading skills through an educational game. To accomplish the robot's objective, we proposed and implemented a lip-syncing system on the APO social robot. The proposed robot's potential with regard to its primary goals, tutoring and practicing lip-reading, was examined through two main experiments. The first experiment was dedicated to evaluating the clarity of the utterances articulated by the robot. The evaluation was quantified by comparing the robot's articulation of words with a video of a human teacher lip-syncing the same words. In this inspection, due to the adults' advanced skill in lip-reading compared to children, twenty-one adult participants were asked to identify the words lip-synced in the two scenarios (the articulation of the robot and the video recorded from the human teacher). Subsequently, the number of words that participants correctly recognized from the robot and the human teacher articulations was considered a metric to evaluate the caliber of the designed lip-syncing system. The outcome of this experiment revealed that no significant differences were observed between the participants' recognition of the robot and the human tutor's articulation of multisyllabic words. Following the validation of the proposed articulatory system, the acceptance of the robot by a group of hearing-impaired participants, eighteen adults and sixteen children, was scrutinized in the second experiment. The adults and the children were asked to fill in two standard questionnaires, UTAUT and SAM, respectively. Our findings revealed that the robot acquired higher scores than the lip-syncing video in most of the questionnaires' items, which could be interpreted as a greater intention of utilizing the APO robot as an assistive technology for lip-reading instruction among adults and children.

Keywords Socially assistive robots · Hearing-impaired · Lip-syncing · Technology acceptance

1 Introduction

In contrast to the previous prevalent concept that considered speech processing an absolute auditory phenomenon, several studies have revealed that visual speech information is complementary to encoding the speech signal due to the multimodal characteristic of speech perception [1, 2]. For example, McGurk and MacDonald demonstrated that discrepancies between the auditory signals that individuals hear and the speakers' visual articulation information could adversely affect auditors' perception concerning the sound they have heard [3]. Accordingly, the audio and visual information acquired concurrently from analyzing speech and focusing on the speaker's articulatory movements leads to

✉ Ali F. Meghdari
meghdari@sharif.edu

✉ Alireza Taheri
artaheri@sharif.edu

[1] Social and Cognitive Robotics Laboratory, Center of Excellence in Design, Robotics, and Automation (CEDRA), Sharif University of Technology, Tehran, Iran

[2] Fereshtegaan International Branch, Chancellor, Islamic Azad University, Tehran, Iran

[3] Department of Humanities, West Tehran Branch, Islamic Azad University, Tehran, Iran

a fuller comprehension of the speech signals [4, 5]. This concept is in accordance with the preference of infants to concentrate on speech synchronized with articulatory gestures rather than improperly linked audio-visual compounds [6]. Hence, lip-reading, which is concerned with eliciting speech information from the movements of jaws, lips, tongue, and teeth within the articulation [7], enhances human speech perception [8]. Although lip-reading capability is considered a complementary skill to improve the auditory perception of hearing people, it plays a vital role throughout hearing-impaired communications [9, 10]. Additionally, the lack of this expertise restricts deaf people's interlocutors solely to individuals who are accustomed to non-verbal communication channels such as Sign Language and Cued Speech. This issue leads deaf people to encounter significant barriers to various two-way communication activities such as university studies [11]. Hence, the accomplishment of deaf people's independence throughout communication is subject to honing their lip-reading skills.

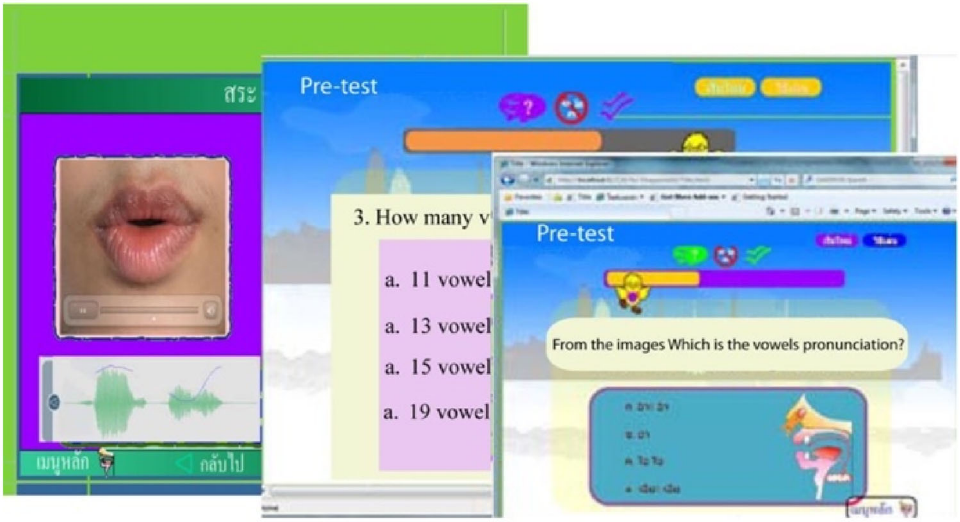
Lip-reading is intrinsically an arduous task. Furthermore, the resemblance between the articulatory elements' activity while pronouncing some letters, such as /g/ and /k/, which are not evident on the lips, hinders the lip reading procedure [12]. Easton and Basala conducted two experiments to assess participants' speech recognition capability from facial gestures. Their examination showed that hearing-impaired observers could only attain an accuracy of 17 ± 12% for 30 one-syllable words and 21 ± 11% for 30 multi-syllable words [13]. Therefore, lip-reading is a challenging task that requires training and practice. Generally, lip-reading teaching approaches can be classified into two main categories, analytic and synthetic methods, which are sometimes deployed cooperatively. In analytical strategies, trainers are asked to analyze the speech's constituents separately, while synthetic methods require participants to synthesize the message according to all presented clues [14].

Several studies have been carried out to determine effective educational programs and investigate the potential effects of the presented methods on learners' achievements in lip reading. Creating instructional videos [14, 15] and developing various computer-based training, test software, and games [16, 17] are well-established techniques in this field. Most of the proposed training programs are comprised of pre-recorded sequences from tutors' articulatory components' activities while pronouncing letters, words, and sentences. Figure 1 depicts the content of a computer game developed for lip-reading instruction.

Thanks to current demographic trends, economic factors, and the growing desire for inclusive education, a flourishing demand for various educational assistive technologies has been generated. Furthermore, literature has indicated that social interaction enhances educational achievements [18, 19]. These premises and the evolution of robotic technology have led to the development of a new concept concerned with utilizing robots that socially interface with humans as teacher's assistants [20]. Several studies conducted in the multidisciplinary field of Human–Robot Interaction (HRI) have highlighted that during interaction with physical robots, individuals' perception and engagement levels are higher than interacting with virtual agents (on-screen characters) [21–23]. Kid and Breazeal explored and compared individuals' responses to a robotic character, an animated character, and a human to examine the impacts of the robot's presence on users' perceptions. Their study revealed that the physical robot is perceived to be more credible, informative, engaging, and enjoyable to interact with compared to the animated character. Two factors of robots, physical presence and real entity, could be credited for the higher effectiveness of physical robots on users' perception in educational scenarios compared to fictional animated characters [24]. Leyzberg et al. scrutinized the influences of the robot's presence on individuals' learning achievements through a robot tutoring task. The outcome of this investigation revealed that a physically-present robot leads to higher learning gains than on-screen characters and enhances participants' performance during the activity [25]. In addition, the robot's embodiment makes people more likely to follow its commands and heightens its authority [26, 27]. As well as the potential positive impacts associated with the physical presence of social robots in education, augmenting humanoid features on social robots that do not necessarily possess a fully human-like appearance improves Human–Robot Interaction [28].

The utility of socially assistive robots in the domain of education has been the subject of several studies performed in the field of HRI [29–31]. The outcomes of these examinations illustrated that social robots' employment within several training subjects, including English as a Foreign Language (EFL) [32–34], Mathematics [35–37], Physics [38], and Programming [39, 40], can be beneficial and enhance the learners' educational gains by keeping the participants engaged [41]. The remarkable results of social robots' utility in the context of education could be extrapolated to teaching cognitive skills to individuals with special needs [42, 43]. Taheri et al. examined the efficacy of the NAO social robot in teaching music to children who were diagnosed with Autism Spectrum Disorder (ASD). The findings of this study revealed that the utility of the social robot as an assistive tool alongside the teacher throughout educational interventions considerably improves learners' performance in information attainment compared with scenarios without the robot [44, 45].

Due to the encouraging prospects of deploying social robots within diverse training programs, in this study, we endeavor to make use of this promising educational technology to enhance hearing-impaired individuals' lip-reading



3. How many v

a. 11 vowel

a. 13 vowel

a. 15 vowel

a. 19 vowel

From the images Which is the vowels pronunciation?

Fig. 1 The environment of a developed lip-reading training software [16]

skills. As previously mentioned, hearing-impaired people struggle to master speech reading skills because of the ambiguity of visual speech cues. This issue, compounded with the lack of attractive instructional tools, makes the training process tedious. Therefore, due to the aforementioned merits of physically-present robots compared to animated characters with respect to improving students' learning performance [25, 46], in this study, an educational assistive robotic platform with capabilities that suit lip-reading training programs was designed to lessen the monotony of the instructional procedure by maintaining participants' level of engagement during the educational scenario.

One of the fundamental issues that should be considered in designing a novel robotic platform is its cost-efficiency, making the robot affordable for its users. Furthermore, the robot should possess assorted complex attributes and dynamic features to maintain participants' engagement throughout long-term interaction [47]. Thus, there is a trade-off between the cost-efficiency aspect and the robot's capacity to perform diverse behaviors. These criteria, alongside the robot's capabilities in terms of accomplishing the desired purposes, determine the appropriate type of robotic head and the required degrees of freedom that should be considered to achieve an effectual design.

The social robot, APO, designed in this study is a tablet-face non-humanoid robot with two degrees of freedom intended to attract and keep learners' attention throughout training programs by performing simple motions. The proposed robot's main features include being cost-effective, easily portable, having a cute design, and possessing an LCD screen to perform lip-reading exercises during interaction with hearing-impaired individuals. A precise lip-syncing system with the capacity to lip-sync any utterance without being formerly registered was also designed and implemented on the robot to achieve APO's instructional purposes and develop a human-like articulation.

The current study comprises two experiments: evaluating the proposed lip-syncing system, which determines the robot's capacity to be employed in educational scenarios, and investigating the acceptance of the APO, which affects participants' cognitive performance and compliance behaviors during interaction with the robot [48]. In the first experiment, a video of a normal-hearing individual pronouncing a set of words was first given to a group of adults, who were asked to watch, identify, and write the pronounced words. Afterward, to assess the robot's visual articulation performance, the participants were asked to sit in front of the APO robot while it articulated the same set of words and then write the words they recognized from the robot's articulation. The success of the lip-syncing system was measured by comparing the participants' perceptions of the lip-synced words in the two scenarios. In this examination, children were excluded from the research participants because they are generally less competent in lip-reading than adults. In the second experiment, the APO robot with the approved lip-syncing capability was involved in interventions to compare the acceptance of two lip-reading training tools (the proposed robotic platform and the silently recorded video). This experiment utilized the Unified Theory of Acceptance and Use of Technology (UTAUT) [49] and the Self-Assessment Manikin (SAM) questionnaires [50] to investigate the robot's acceptability to hearing-impaired adults and children, respectively. Figure 2 depicts the study's contents schematically.

The rest of this paper is organized as follows. Section 2 is dedicated to various aspects of designing a new domestic robotic platform, the APO social robot. In Sect. 3, the proposed lip-syncing system is delineated. Section 4 addresses the approach adopted to assess the developed lip-syncing system. The intervention scenario arranged to examine the acceptance of the robot is also covered in this section. In Sects. 5 and 6, the experiments' results and discussion are

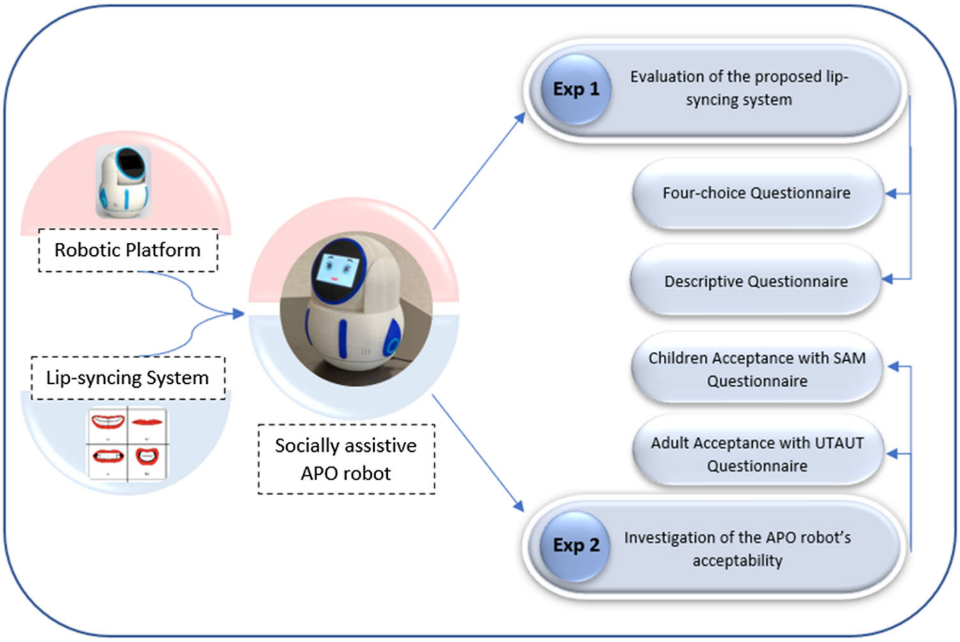


Fig. 2 The schematic contents of the study

presented, respectively. Section 7 is devoted to the limitations of the current study. Finally, in Sect. 8, the concluding remarks are declared.

2 The "APO" Robotic Platform

2.1 Conceptual Design

Providing individuals with a high-quality education is one of the principal concerns of developed societies. The COVID-19 pandemic not only led to burgeoning demands for remote training programs but also demonstrated the necessity of developing assistive technologies. The use of social robots as assistive tools for educational applications has attracted growing attention among researchers due to their potential to engage students and enhance their learning achievements [51]. This section is dedicated to designing a simple non-humanoid social robot named APO for educational purposes. The primary objective of developing the APO robot was to perform the role of hearing-impaired individuals' playmates through a lip-reading educational game. Nevertheless, APO could be utilized within wide-ranging tutoring applications outside of the original intent. Cuteness, cost-effectiveness, mobility, and practicality were the main factors in the APO robot's design. Figures 3 and 4 depict simple initial sketches and the conceptual design of the APO robot, respectively.

2.2 APO robot's Hardware Design

The APO robot's platform comprises three primary compounds, a tablet-face head, the robot's upper body, and lower body parts. Figure 5 demonstrates the APO robotic platform's final 3D model and exploded view.

The robot includes two rotary degrees of freedom (DOFs); the first one concerns the relative rotation between the robot's upper and lower parts with respect to the roll axis, and the second relates to the pitch rotation of the robot's head, stated in relation to the upper body part. Hence, provided that the APO's lower body is fixed, the robot is capable of looking at any desired points located in the robot's surroundings. The proposed robotic platform is equipped with a camera and a microphone, considered as the audio and visual input devices, as well as a 5-inch tablet face and speakers, regarded as output hardware that enhances the robot's capabilities to conduct both verbal and non-verbal communications with users. Moreover, a Raspberry Pi computer performs the robot's internal processing. Table 1 summarizes the robot's features.

2.3 APO robot's Software Design

A Graphical User Interface (GUI) was developed for the APO robot to customize the robot's operational system to be controllable by non-expert users. Through the designed GUI, the users would be able to control the APO robot's motions, stream its camera on computers or tablets, change the robot's facial expressions, vary the color and the brightness of LEDs, and ultimately type any utterance to be lip-synced by the robot. Figure 6 depicts the environment of the GUI developed for the APO robot.

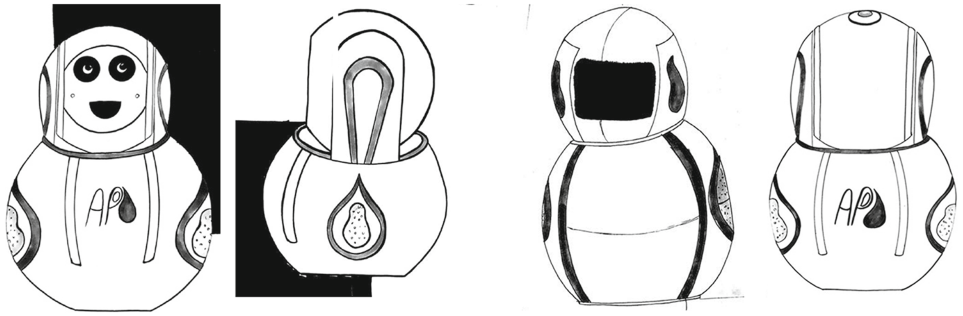


Fig. 3 The simple initial sketches of the APO robot

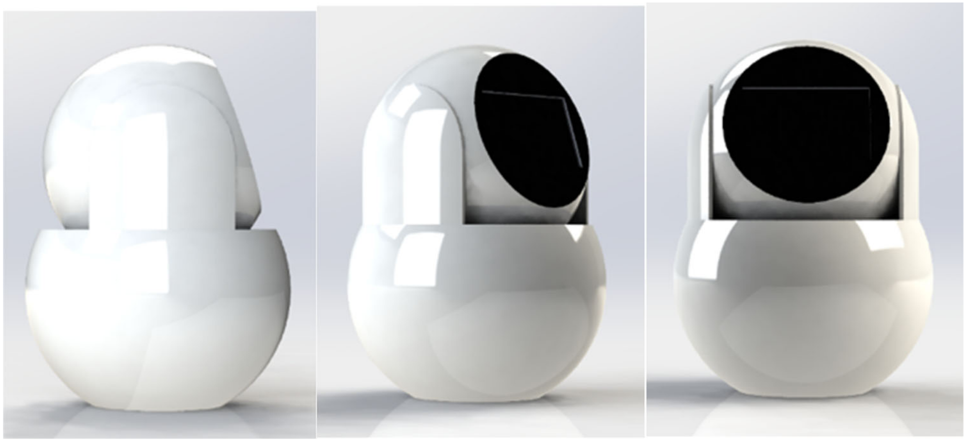


Fig. 4 The conceptual design of the APO robot

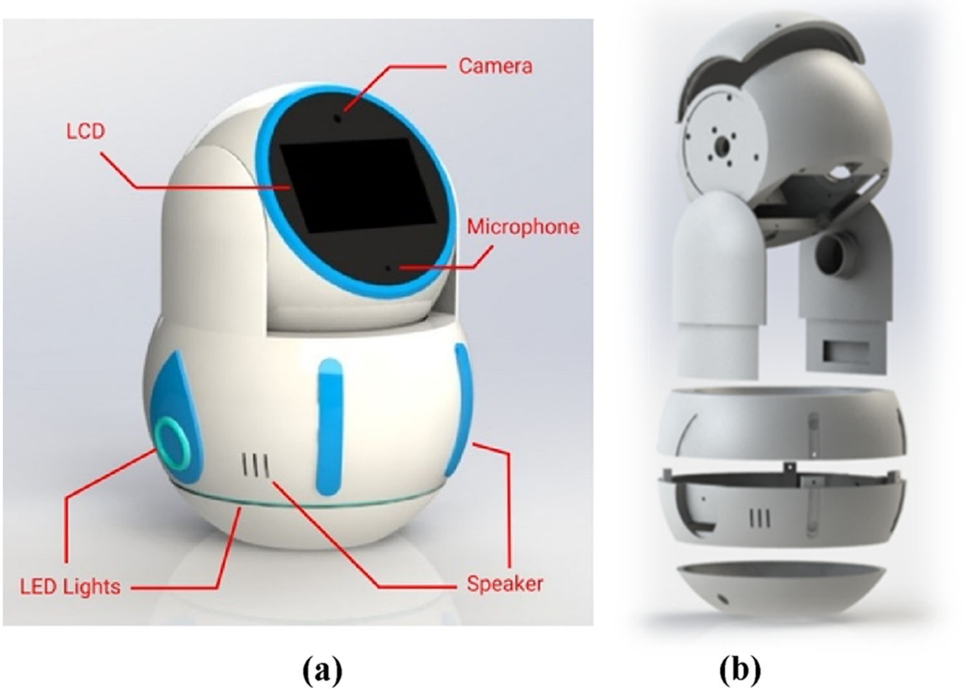


Fig. 5 **a** The APO robot's 3D and **b** exploded view

Table 1 The specifications of the APO social robot

Dimension	23 × 23 × 21.2 (cm^3)
Weight	2.7 kg
Actuators	2 Servomotors
Sensors	Camera (Raspberry Pi Camera Module 2) Microphone
DOFs	2 degrees of freedom
Operating section	Raspberry Pi 3
Operating system	ROS on Ubuntu 15
Power section	12 DC Voltage

3 Lip-syncing System

3.1 Lip-syncing Shapes Design

Adding the lip-syncing capability to a social robot may augment users' perception of the robot's verbal communication and enhance the potential for tutoring word pronunciation. This characteristic becomes even more crucial when the robot is interacts with hearing-impaired individuals. Hence, a realistic design of articulatory visual elements is consequential. In an effort to produce the most sensible lip shape design, an Iranian Sign Language (ISL) interpreter was hired to exaggerate the pronunciation of the letters of the alphabet. While he was pronouncing each letter (including vowels and consonants), several images were captured in a straight-ahead position. The most detailed image of each letter was used to design the visual parts of the robot's mouth. Figure 7 shows the design procedure performed for each letter.

Individual designs were produced in this manner for all letters of the alphabet, including consonants and vowels. Figure 8 demonstrates the designed alphabet shapes.

3.2 Lip Morphing

Developing a dictionary comprised of several sequences for each word is a lengthy process and requires massive computer memory. Hence, developing an algorithm that can receive a word as an input, disassemble it into its constituent letters, and then smoothly morph the letters into each other would be a more efficient way. A principal consideration in developing an algorithm that will morph the mouth's elements into their equivalents in successive frames to attain natural verbal communication is that the deformation of the mouth's features should be minimized.

The transition problem is daunting in the animation field [52]. It should be noted that although a spectator can disregard flaws in drawing and schematic elements, unnatural or discrete motions are not permissible [53]. The change between the initial and final point forms the fluidity of a movement, and following the path between the initial and final points in a linear manner leads to an unnatural transition [53]. Adding acceleration terms produces a dynamic tween that enhances the transition's degree of goodness [54]. In animation jargon, easing is equivalent to morphing, which is a combination of several tools used to specify the manner in which elements transition into their corresponding elements in consecutive frames. An Easing Function is a function that determines the way that the transition from the initial point to the final point occurs with respect to terms of velocity and acceleration. Several functions can be employed to fulfill this purpose. Figure 9 shows some of these functions.

Following an investigation of the diverse easing functions shown in Fig. 9, the InOutExpo function was chosen due to its ability to naturally and smoothly transition the elements of the robot's mouth. As Fig. 9 depicts, the velocity at the beginning and the end of the time interval is zero in the InOutExpo

Fig. 6 The APO robot's GUI

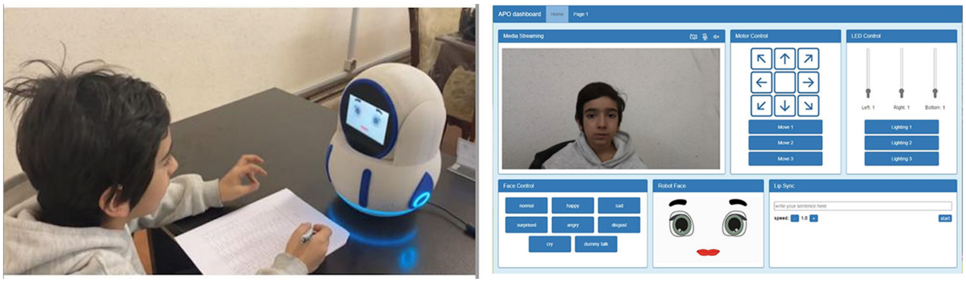


Fig. 7 The procedure of designing the robot's lips while pronouncing various letters

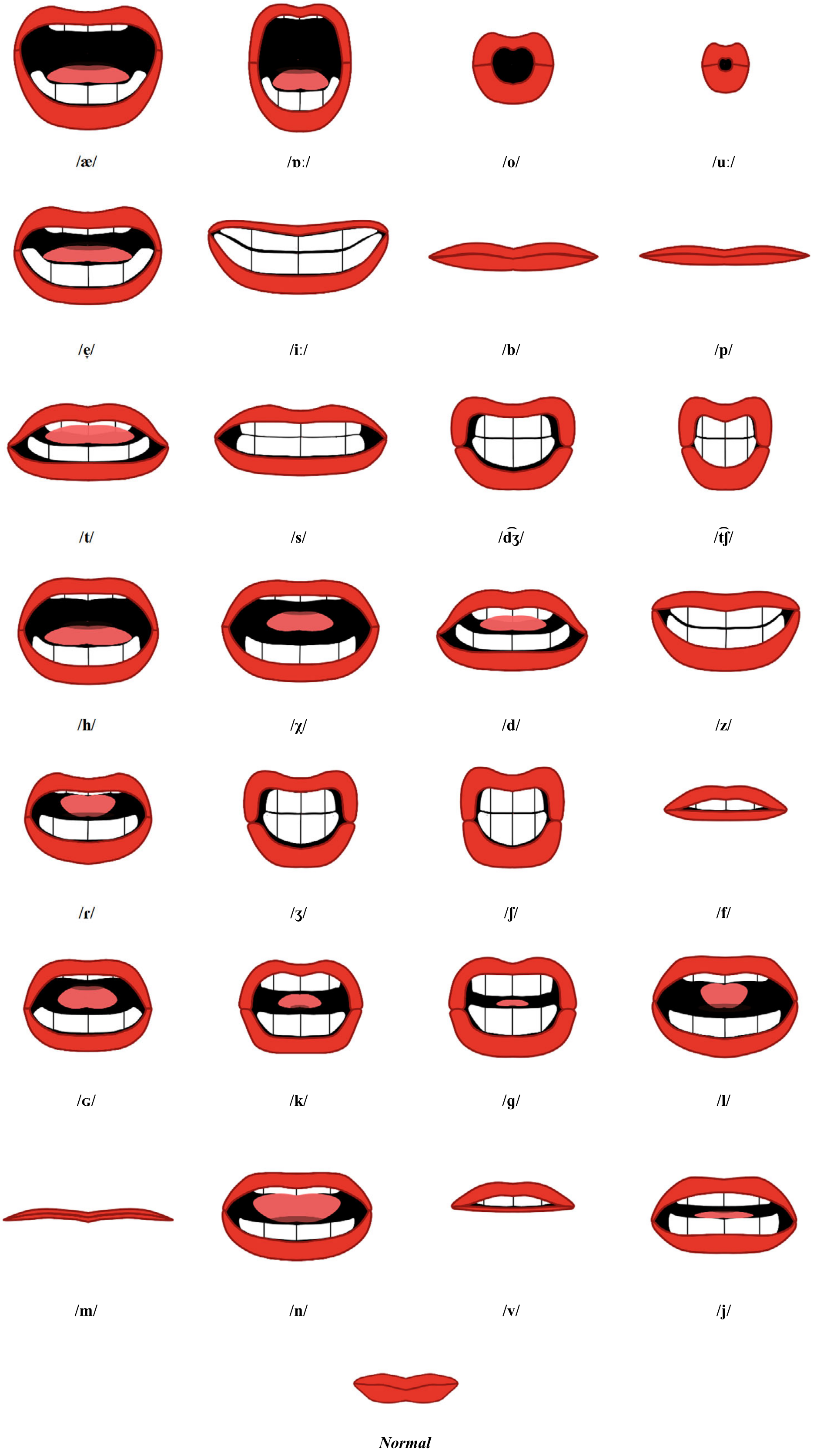


Fig. 8 The designed lip shapes for each letter of the Persian alphabet

easing function, which leads to an aesthetically pleasing transition. The equation of this function is as follows [53]:

$$y = \begin{cases} 0 & x = 0 \\ 2^{20x-11} & x \in \left(0, \frac{1}{2}\right] \\ 1 - 2^{-20x+9} & x \in \left(\frac{1}{2}, 1\right) \\ 1 & x = 1 \end{cases} \tag{1}$$

As Fig. 8 depicts, the mouth's elements are fundamentally composed of curved lines that form closed curves. An operational method of making the transition from one mouth state to another within successive frames is to divide each curve into numerous points and utilize the easing function described in Eq. (1) for each point to smooth the transition process. The subsequent challenge is to find corresponding points between two shapes within successive frames. Achieving a natural form of speech is subject to minimizing the articulators' deformation; consequently, corresponding points should be chosen to minimize the sum of the tracks followed by the points during the transition. The penalty function that describes this issue is as follows:

$$J = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \widehat{x}_i)^2}{N^2}} \tag{2}$$

Thus, the transition problem is simplified to minimize the above cost function. The more the number of chosen points increases, the better the transition smoothness is and the greater the computational cost accrues.

Ordinarily, the term easing alludes to animation made for games and HTML applications. Qt and jQuery libraries are also utilized to implement transition functions. This study used JavaScript and HTML and benefitted from the KUTE library to implement the transition function. Also, Adobe Illustrator was used to draw the articulators' scheme. The developed module takes any word, phrase, sentence, and time parameter as inputs and performs the lip-syncing in that time. The nature of the transition was assessed by a visual examination by the sign language interpreter who cooperated with our research group. Figure 10 illustrates the way that the developed algorithm executes.

4 Methodology

The current study is composed of two principal experiments. The first experiment evaluates the developed lip-syncing system by comparing the participants' perception of a set of words articulated by the robot and a human tutor, and the second experiment investigates the acceptance of the robot through interaction with adults and children.

4.1 Participants

To appraise the explicitness of the robot's visual articulation performance, a group of Iranian adults studying at Fereshtegan International Branch of the Islamic Azad University and skilled in lip-reading and sign language took part in the first experiment. The under-investigation group was composed of seven deaf individuals (three men, four women, mean age = 20.43, standard deviation = 1.72), seven hard-of-hearing persons (four men, three women, mean age = 20.29, standard deviation = 1.80), and seven normal hearing students (four men, three women, mean age = 20.14, standard deviation = 1.35). Another group of 34 Iranian hearing-impaired individuals participated in the second experiment. This group, composed of 16 children (nine boys, seven girls, mean age = 7.63, standard deviation = 1.15) and 18 adults (ten men, eight women, mean age = 20.78, standard deviation = 1.87), helped explore the acceptance of the APO robot through interaction with its target groups. All participants had no previous experience interacting with social robots.

4.2 Assessment Tools

In the first experiment, two types of tests, descriptive and four-choice, were adopted. Both tests were comprised of twenty-eight questions concerned with the lip-synced words. Throughout the descriptive test, participants were asked to guess and write the articulated words, while in the four-choice test, they were required to choose the correct answer among four semi-syllable words. The collection of words was composed of fourteen one-syllable and fourteen multi-syllable words. Each of the sounds (consonants and vowels) was repeated more than three times throughout the set. In both trials, normalized scores, defined as the ratio of the participants' correct responses to the total number of questions, were utilized as metrics to compare the perceptibility of the utterances lip-synced by the robot and the human teacher.

Furthermore, the UTAUT [55] and SAM [50] questionnaires were utilized to evaluate the adults' and children's acceptance of the APO robot, respectively. The UTAUT questionnaire is a well-established test developed to quantify the acceptance of technology by older adults. This model aims to assess the users' intention to employ a novel technology according to performance expectancy, effort expectancy, social influence, and facilitating conditions. Individual attributes, including gender, age, experience, and voluntariness, are the main moderating influences [56]. The UTAUT model requires some modification to fit the field of the technology it is applied in; therefore, we employed the version modified by Heerink et al., which is suitable in the social robotics context [49]. This modified version investigates twelve constructs, including Anxiety (ANX),

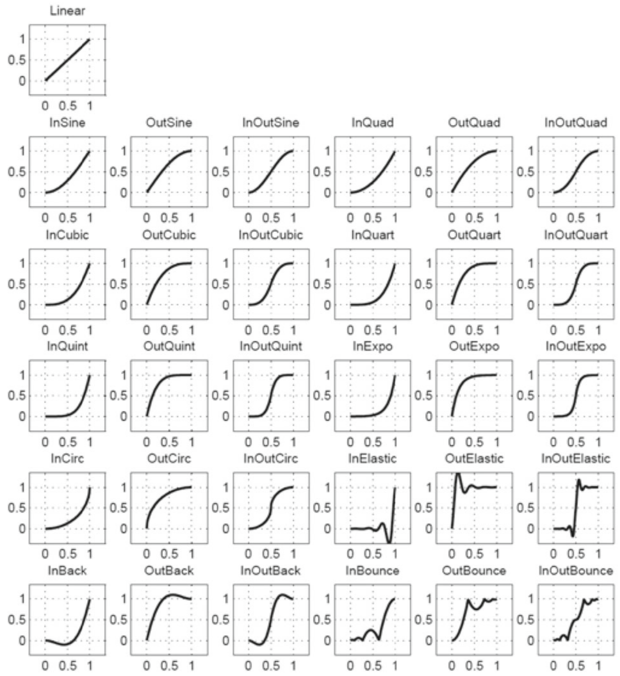


Fig. 9 Penner's easing functions [53]

Attitude Towards Technology (ATT), Facilitating Conditions (FC), Intention to Use (ITU), Perceived Adaptiveness (PAD), Perceived Enjoyment (PENJ), Perceived Ease of Use (PEOU), Perceived Sociability (PS), Perceived Usefulness (PU), Social Influence (SI), Social Presence (SP), and Trust. The second questionnaire, i.e., the SAM model, was used to assess the robot's acceptability by children. This test measures pleasure, arousal, and dominance concerned with the social robot [50]. Both questionnaires, UTAUT and SAM, were scored via Five-Likert pictorial scales (range:1–5) [48, 57].

4.3 Procedure

4.3.1 Experiment one: Evaluation of the Proposed Lip-Syncing System

In the study's first phase, an educational lip-reading game was designed to evaluate the proposed lip-syncing system's capacity for conveying messages to hearing-impaired people. The objective of this game was to compare the APO robot with human articulations. To this end, we asked a tutor working at the Fereshtegan University to lip-sync a set of Persian words and recorded a video while she was articulating them. The robotic platform also lip-synced the same set of words in a different succession. Subsequently, the adult participants were asked to watch the silent sequence and specify the terms they had recognized due to their greater expertise in lip-reading compared to children. Then, they were asked to take the words perception test with the APO robot in the same manner. Afterward, the students were scored according to the number of words correctly identified in each of the two lip-reading scenarios. The lip-reading games were carried out twice to provide the participants with clues about the correct answers. The first time, they were asked to note down the words they saw, while the second time, they were required to choose the correct answers from a four-choice test. Furthermore, the counterbalancing technique was used to eliminate the order effect. In this regard, half of the participants first encountered the robot and then the lip-reading video, while others underwent the reverse order.

4.3.2 Experiment Two: Investigation of the Robot's Acceptability

The second phase of the study was devoted to comparing the acceptability of the two lip-reading training programs, the lip-reading tutoring program taught by the APO robot and the one performed by the video of the tutor while she was lip-syncing the same words. The investigation was conducted on both adults and children. In this regard, first, the adult participants were asked to complete the UTAUT questionnaire. Afterward, the acceptability test was performed through interaction with the children. The SAM questionnaire was employed in this examination to quantify the children's emotional responses. Following the meeting and playing with the APO robot. Figure 11 depicts the experimental setup in the two experiments.

5 Results

5.1 Experiment one: Evaluation of the Proposed lip-syncing System

The first experiment was dedicated to comparing the participants' perception of the APO robot's visual articulation and the video of the tutor lip-syncing the same set of words. In

Fig. 10 The morphing algorithm demo while pronouncing (/ɒːb/), which means water in Persian

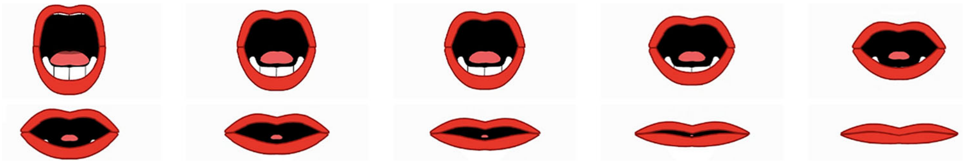


Fig. 11 **a** Experiment one: Evaluation of the proposed lip-syncing system, **b** Experiment two: Investigation of the robot's acceptability

the first step of the experiment, the participants were asked to write the utterances they had recognized in the two lip-reading games, while in the second step, they were asked to choose correct answers from a four-choice questionnaire. To check the quality of the robot's articulatory system, the participants' normalized scores in the two lip-reading scenarios were first statistically analyzed (t-test) using Minitab software. Then, the resulting p-values were utilized as metrics to determine whether participants' perceptions of the two lip-reading games had significant differences.

Furthermore, a similarity score that states the number of words identically labeled (correctly or incorrectly) in the lip-reading games was also defined to measure the comparability of the robot's lip-syncing system and the human articulation. In other words, the similarity score explains the number of words similarly identified in the two games, regardless of their correctness, normalized by the total words. Table 2 summarized the participants' normalized scores in this experiment.

5.2 Experiment Two: Investigation of the Robot's Acceptability

In the second experiment, the acceptance of the APO robot among adults and children was investigated in comparison with the recorded lip-syncing video throughout the designed lip-reading games. Following the end of the two games, the children were asked about their preference between the two educational games to investigate the Child-Robot interaction. The SAM questionnaire was given to the children following the end of the two games, and they were asked to answer each item via Likert scores ranging from one to five. The participants' scores concerning the robot and the recorded video, as well as the statistical analysis (t-test) results, are summarized in Table 3.

To scrutinize the robot's acceptability to the adult participants, the UTAUT questionnaire was employed. Like the previous procedure, the adults were asked to declare their preference for the APO robot and the recorded video following the two games. Afterward, the participants were asked to complete the UTAUT test via the Five-Likert scale. To probe into the differences between the acceptance of the robot and the recorded video, statistical analysis was performed. Table 4 presents the results concerned with the robot and the video acceptance scores.

6 Discussion

6.1 Experiment One: Evaluation of the Proposed Lip-syncing System

Throughout the first experiment, certain words were correctly recognized in the human tutor's video and the APO robot lip-syncing, while other items were misunderstood in lip-reading games. Interestingly, most of these misconceptions about the articulated words were the same. The statistical analysis (t-test) of the above results (for an alpha of 0.05) revealed that the scores corresponding to the participants' recognition of monosyllabic words in the lip-syncing video were significantly higher than the APO robot's articulation ($p < 0.05$). However, their comprehension of multisyllabic words was not significantly different in the two games ($p > 0.05$). Therefore, we concluded that the robot's articulation of words composed of two or more syllables is comparable with the human tutor. However, a power analysis using G*Power 3.1 Software [58] revealed that determining a statistically significant difference requires at least $N = 35$ participants for this experiment based on the medium effect size of 0.5, a power level of 0.8, and a significance level of 0.05. Hence, our findings, with respect to the comparability of the robot and the human tutor articulation of multisyllabic words, should be reported cautiously due to the limited number of participants. Additionally, the preliminary exploratory findings of the first experiment demonstrated that the hard-of-hearing participants were more competent at both lip-reading games than the other groups.

Table 2 The participants' scores in the two lip-reading scenarios (the APO robot and the human tutor)

Group	Items	Test type	Score's mean (SD)		Similarity	*p*-value
			Robot	Video		
Deaf group	Monosyllabic	Descriptive	0.3163 (0.0382)	0.4184 (0.1046)	0.4538	0.032
		Four-Choice	0.5408 (0.0998)	0.6531 (0.0764)	0.5936	0.036
	Multisyllabic	Descriptive	0.5816 (0.0961)	0.6327 (0.0643)	0.5714	0.266
		Four-Choice	0.6735 (0.0382)	0.7143 (0.0583)	0.7032	0.147
Hard hearing group	Monosyllabic	Descriptive	0.3571 (0.0825)	0.4694 (0.1080)	0.5128	0.049
		Four-Choice	0.6020 (0.0998)	0.7245 (0.0764)	0.7593	0.024
	Multisyllabic	Descriptive	0.6531 (0.0961)	0.6735 (0.0697)	0.7863	0.657
		Four-Choice	0.7449 (0.0909)	0.7755 (0.0868)	0.8542	0.531
Normal-hearing group	Monosyllabic	Descriptive	0.2857 (0.0922)	0.3878 (0.0810)	0.4126	0.048
		Four-Choice	0.4796 (0.1069)	0.6020 (0.0818)	0.4921	0.033
	Multisyllabic	Descriptive	0.5306 (0.0697)	0.6020 (0.0909)	0.5574	0.125
		Four-Choice	0.6429 (0.0583)	0.6939 (0.0795)	0.6828	0.196

Table 3 The SAM test results and analysis comparing the children's acceptance of the APO robot and the recorded video

Item	Score's mean (SD)		*p*-value
	Robot	Video	
Pleasure	**4.467 (0.443)**	**3.883 (0.472)**	**0.011**
Arousal	**3.783 (0.846)**	**3.100 (0.545)**	**0.046**
Dominance	3.867 (0.864)	3.300 (0.706)	0.126

Table 4 The UTAUT test results and analysis comparing the adults' acceptance of the APO robot and the recorded video

Item	Score's mean (SD)		*p*-value
	Robot	Video	
ANX	**2.700 (1.160)**	**4.000 (0.816)**	**0.010**
ATT	**4.200 (0.789)**	**2.900 (1.101)**	**0.007**
FC	3.400 (0.966)	3.100 (0.994)	0.503
ITU	**4.100 (0.876)**	**3.100 (0.994)**	**0.028**
PAD	**3.500 (0.850)**	**2.400 (0.966)**	**0.015**
PENJ	**4.400 (0.699)**	**3.100 (0.738)**	**0.001**
PEOU	3.100 (0.994)	3.700 (1.059)	0.208
PS	**4.200 (0.919)**	**3.000 (1.333)**	**0.031**
PU	4.000 (0.816)	3.900 (0.876)	0.795
SI	**3.800 (1.033)**	**2.700 (1.160)**	**0.038**
SP	3.600 (0.966)	2.700 (0.949)	0.050
Trust	3.700 (0.949)	3.800 (1.033)	0.824

6.2 Experiment Two: Investigation of the Robot's Acceptability

A review of the children's acceptance of the robot during their interaction in the study showed that the APO achieved higher scores than the recorded video on all items of the SAM questionnaire. Additionally, according to the statistical analysis (for an alpha of 0.05), significant differences were observed for the metrics concerned with pleasure and arousal in the two scenarios ($p < 0.05$), while the dominance item showed no significant difference between the two games ($p > 0.05$). The acceptance of the robot among adults, measured via the UTAUT questionnaire, indicated that the average scores of the APO robot for the ATT, FC, ITU, PAD, PENJ, PS, PU, SI, and SP items, were higher than the recorded video, but significant differences were only observed ($p < 0.05$) for the ATT, ITU, PAD, PENJ, PS, SI, and SP constructs. The robot's significantly higher scores in the PAD and PENJ items are due to the APO robot's enjoyability and easiness of use. The SP and PS constructs correspond to the robotic platform's adaptability and sociability, leading to a positive image of the APO characteristics. Higher scores on the ATT and ITU items show a higher engagement level for the social robot compared to the video throughout the lip-reading training procedure. The significant difference in the SI item describes the participants' preference to share this technology with others. The outcome of this examination should be reported with caution due to the limited number of participants calculated by the conducted power analysis.

7 Limitations and Future Work

The robotic platform developed in the current study utilized an LCD screen to lip-sync the words. One of the robot's limitations is the 2D trait of the proposed lip-syncing system. This issue complicates the participants' perception of the APO robot's utterances. Another limitation of our study was the number of volunteers. Due to the COVID-19 pandemic, only a small number of people agreed to take part in our examination. Our future work objective is to enhance the lip-syncing system to be more analogous to the articulation of a human tutor. Additionally, future studies will focus on designing an attractive collaborative game to increase hearing-impaired people's lip-reading skills by developing a lip-reading system using CNN and RNN models to improve the lip-reading capability of the APO robot. In this way, the robot will be regarded as the playmate of learners playing the interactive lip-reading game. This interaction engages individuals through learning procedures and assesses their performance while lip-syncing the target words. Furthermore, comparing the lip-reading attainments through Robot-Assisted Therapy (RAT) sessions and conventional lip-reading instructions is another possible subject for future experiments.

8 Conclusion

This study proposed a new tablet-face robotic platform, APO, that benefits from a lightweight and portable platform designed to enhance lip-reading training programs for hearing-impaired individuals. In this regard, a lip-syncing system based on the visual articulation of a sign language interpreter was developed and implemented on the robot to accomplish the robot's educational objective. To assess the efficacy of the developed robotic platform's desired objective, two main experiments were conducted, evaluating the proposed lip-syncing system and investigating the acceptance of the robot among children and adults. The first experiment's analysis indicated that the proposed lip-syncing system performed appropriately regarding the articulation of compound words. Moreover, the exploratory outcomes of the investigation revealed that hard-of-hearing participants were more adept at comprehending lip-synced words. The outcome of the second experiment, the examination of the acceptance of the robot among both adults and children, also demonstrated that the robot scored higher acceptance than a recorded video when employed as assistive technology for a lip-reading training program. However, the reported outcomes of this study should be interpreted cautiously due to the small number of subjects, as estimated by the power analysis.

Acknowledgements This study was funded by the "Dr. Ali Akbar Siassi Memorial Research Grant Award" and Sharif University of Technology (Grant No. G980517). We also thank Mrs. Shari Holderread for the English editing of the final manuscript.

Authors' contributions All authors contributed to the study's conception and design. Material preparation, data collection, and analysis were performed by Alireza Esfandbod, Ahmad Nourbala, and Zeynab Rokhi. The first draft of the manuscript was written by Alireza Esfandbod and Zeynab Rokhi. All authors read and approved the final manuscript.

Data availability All data from this project (videos of the sessions, results of the questionnaires, scores of performances, etc.) are available in the archive of the Social & Cognitive Robotics Laboratory.

Code availability All of the codes are available in the archive of the Social & Cognitive Robotics Laboratory. In case the readers need the codes, they may contact the corresponding author.

Declarations

Conflict of interest Author Alireza Taheri has received a research grant from the Sharif University of Technology (Grant No. G980517). The authors Alireza Esfandbod, Ahmad Nourbala, Zeynab Rokhi, Ali F. Meghdari, and Minoo Alemi assert that they have no conflict of interest.

Ethical approval Ethical approval for the protocol of this study was provided by the Iran University of Medical Sciences (#IR.IUMS.REC.1395.95301469).

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication The authors affirm that human research participants provided informed consent for the publication of the image in Figs. 6 and 11. All of the participants have consented to the submission of the results of this study to the journal.

References

1. Dodd B (1979) Lip reading in infants: Attention to speech presented in-and out-of-synchrony. Cogn Psychol 11(4):478–484
2. L. D. Rosenblum (2008) Primacy of multimodal speech perception
3. McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264(5588):746–748
4. Sumby WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. J Acoust Soc Am 26(2):212–215
5. Erber NP (1975) Auditory-visual perception of speech. J Speech Hearing Disorders 40(4):481–492
6. MacKain K, Studdert-Kennedy M, Spieker S, Stern D (1983) Infant intermodal speech perception is a left-hemisphere function. Science 219(4590):1347–1349
7. Campbell R, Zihl J, Massaro D, Munhall K, Cohen M (1997) Speechreading in the akinetopsic patient LM. Brain A J Neurol 120(10):1793–1803
8. Alegria J, Charlier BL, Mattys S (1999) The role of lip-reading and cued speech in the processing of phonological information in French-educated deaf children. Eur J Cogn Psychol 11(4):451–472
9. Conrad R (1977) Lip-reading by deaf and hearing children. Br J Educ Psychol 47(1):60–65
10. Dodd B (1977) The role of vision in the perception of speech. Perception 6(1):31–40
11. Noble H (2010) Improving the experience of deaf students in higher education. British J Nurs 19(13):851–854
12. Woll B (2012) Speechreading revisited. Deaf Educ Int 14(1):16–21
13. Easton RD, Basala M (1982) Perceptual dominance during lipreading. Percept Psychophys 32(6):562–570
14. Dodd B, Plant G, Gregory M (1989) Teaching lip-reading: the efficacy of lessons on video. Br J Audiol 23(3):229–238

15. Kyle FE, Campbell R, Mohammed T, Coleman M, MacSweeney M (2013) Speechreading development in deaf and hearing children: introducing the test of child speechreading. J Speech Lang Hear Res 56(2):416–426. https://doi.org/10.1044/1092-4388(2012/12-0039)
16. Chaisanit S, Suksakulchai S (2014) The E-learning platform for pronunciation training for the hearing-impaired. Int J Multim Ubiquit Eng 9(8):377–388
17. Nittaya W, Wetchasit K, Silanon K (2018) Thai Lip-Reading CAI for hearing impairment student. In: in *2018* seventh ICT international student project conference*(ICT-ISPC)*, 2018: IEEE, pp. 1–4
18. Gorham J (1988) The relationship between verbal teacher immediacy behaviors and student learning. Commun Educ 37(1):40–53
19. Witt PL, Wheeless LR, Allen M (2004) A meta-analytical review of the relationship between teacher immediacy and student learning. Commun Monogr 71(2):184–207
20. Tanaka F, Matsuzoe S (2012) Children teach a care-receiving robot to promote their learning: field experiments in a classroom for vocabulary learning. J Human-Robot Inter 1(1):78–95
21. Alemi M, Abdollahi A (2021) A cross-cultural investigation on attitudes towards social robots: Iranian and Chinese University students. J Higher Edu Policy Leadership Studies 2(3):120–138
22. Li J (2015) The benefit of being physically present: a survey of experimental works comparing copresent robots, telepresent robots and virtual agents. Int J Hum Comput Stud 77:23–37
23. Wainer J, Feil-Seifer DJ, Shell DA, Mataric MJ, (2007) Embodiment and human-robot interaction: A task-based perspective. In: RO-MAN 2007-The 16th IEEE international symposium on robot and human interactive communication, IEEE, pp. 872–877
24. Kidd CD, Breazeal C. (2004) Effect of a robot on user perceptions. In: 2004 IEEE/RSJ international conference on intelligent robots and systems (IROS)(IEEE Cat. No. 04CH37566), vol. 4: IEEE, pp. 3559–3564
25. Leyzberg D, Spaulding S, Toneva M, Scassellati B. (2012) The physical presence of a robot tutor increases cognitive learning gains. In: Proceedings of the annual meeting of the cognitive science society, vol. 34(34)
26. Bainbridge WA, Hart J, Kim ES, Scassellati B, (2008) The effect of presence on human-robot interaction. In: RO-MAN 2008-The 17th IEEE international symposium on robot and human interactive communication, IEEE, pp. 701–706
27. Bainbridge WA, Hart JW, Kim ES, Scassellati B (2011) The benefits of interactions with physically present robots over video-displayed agents. Int J Soc Robot 3(1):41–52
28. Duffy BR (2003) Anthropomorphism and the social robot. Robot Auton Syst 42(3–4):177–190
29. Belpaeme T, Kennedy J, Ramachandran A, Scassellati B, Tanaka F (2018) Social robots for education: a review. Science Robotics 3(21):eaat5954
30. Tanaka F, Isshiki K, Takahashi F, Uekusa M, Sei R, Hayashi K, (2015) Pepper learns together with children: Development of an educational application. In :2015 IEEE-RAS 15th international conference on humanoid robots (Humanoids), IEEE, pp. 270–275
31. Leite I, Pereira A, Castellano G, Mascarenhas S, Martinho C, Paiva A, (2011) Social robots in learning environments: a case study of an empathic chess companion. In: Proceedings of the international workshop on personalization approaches in learning environments, vol. 732, pp. 8-12
32. Alemi M, Meghdari A, Ghazisaedy M (2014) Employing humanoid robots for teaching English language in Iranian junior high-schools. Int J Humanoid Rob 11(03):1450022
33. Alemi M, Meghdari A, Ghazisaedy M (2015) The impact of social robotics on L2 learners' anxiety and attitude in English vocabulary acquisition. Int J Soc Robot 7(4):523–535
34. Gordon G. et al., (2016) Affective personalization of a social robot tutor for children's second language skills. In: Proceedings of the AAAI conference on artificial intelligence, vol. 30(1)
35. Brown LN, Howard AM (2014), The positive effects of verbal encouragement in mathematics education using a social robot. In: 2014 IEEE integrated STEM education conference, IEEE, pp. 1–5
36. Zhong B, Xia L (2020) A systematic review on exploring the potential of educational robotics in mathematics education. Int J Sci Math Educ 18(1):79–101
37. Reyes GEB, López E, Ponce P, Mazón N (2021) Role assignment analysis of an assistive robotic platform in a high school mathematics class, through a gamification and usability evaluation. Int J Soc Robot 13(5):1063–1078
38. Badeleh A (2021) The effects of robotics training on students' creativity and learning in physics. Educ Inf Technol 26(2):1353–1365
39. Chioccariello A, Manca S, Sarti L (2004) Children's playful learning with a robotic construction kit. Developing New Technologies for young Children, pp. 93–112
40. González YA, Muñoz-Repiso AG (2018) A robotics-based approach to foster programming skills and computational thinking: pilot experience in the classroom of early childhood education. In: Proceedings of the 6th international conference on technological ecosystems for enhancing multiculturality, pp. 41–45
41. Rosanda V, Istenic Starcic A, (2019) The robot in the classroom: a review of a robot role. In: International symposium on emerging technologies for education, Springer pp. 347–357
42. Dautenhahn K et al (2009) KASPAR–a minimally expressive humanoid robot for human–robot interaction research. Appl Bionics Biomech 6(3):369–397
43. Wood LJ, Robins B, Lakatos G, Syrdal DS, Zaraki A, Dautenhahn K (2019) Developing a protocol and experimental setup for using a humanoid robot to assist children with autism to develop visual perspective taking skills. Paladyn, J Behav Robot 10(1):167–179
44. Taheri A, Shariati A, Heidari R, Shahab M, Alemi M, Meghdari A (2021) Impacts of using a social robot to teach music to children with low-functioning autism. Paladyn, J Behav Robot 12(1):256–275
45. Taheri A, Meghdari A, Alemi M, Pouretemad H (2019) Teaching music to children with autism: a social robotics challenge. Scientia Iranica 26:40–58
46. Belpaeme T, Kennedy J, Ramachandran A, Scassellati B, Tanaka F (2018) Social robots for education: a review. Science Robotics 3(21):eaat5954
47. Leite I, Martinho C, Paiva A (2013) Social robots for long-term interaction: a survey. Int J Soc Robot 5(2):291–308
48. Maggi G, Dell'Aquila E, Cucciniello I, Rossi S (2020) Don't get distracted!": the role of social robots' Interaction Style on Users' cognitive performance, acceptance, and non-compliant behavior. Int J of Soc Robotics 13:2057–2069
49. Heerink M, Kröse B, Evers V, Wielinga B (2010) Assessing acceptance of assistive social agent technology by older adults: the almere model. Int J Soc Robot 2(4):361–375
50. Bradley MM, Lang PJ (1994) Measuring emotion: the self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatry 25(1):49–59
51. Ceha J, Law E, Kulić D, Oudeyer P-Y, Roy D (2022) Identifying functions and behaviours of social robots for in-class learning activities: Teachers' perspective. Int J Soc Robot 14(3):747–761
52. Parent R (2012), Computer Animation, 3rd Revised edn. Ed: Morgan Kaufmann, Burlington
53. Izdebski Ł, Sawicki D (2016) Easing functions in the new form based on bézier curves. In: International conference on computer vision and graphics, Springer pp. 37–48
54. Penner R. (2002) Motion, tweening, and easing. *Programming Macromedia Flash MX,* pp. 191–240

55. Venkatesh V, Morris MG, Davis GB, Davis FD (2003) User acceptance of information technology: toward a unified view. MIS Quarterly 27:425–478
56. Venkatesh V, Thong JY, Xu X (2016) Unified theory of acceptance and use of technology: a synthesis and the road ahead. J Assoc Inf Syst 17(5):328–376
57. Striepe H, Donnermann M, Lein M, Lugrin B (2021) Modeling and evaluating emotion, contextual head movement and voices for a social robot storyteller. Int J Soc Robot 13(3):441–457
58. Faul F, Erdfelder E, Buchner A, Lang A-G (2009) Statistical power analyses using G* Power 3.1: tests for correlation and regression analyses. Behav Res Methods 41(4):1149–1160

Alireza Esfandbod is a Ph.D. candidate at the Mechanical Engineering Department of the Sharif University of Technology, Tehran, Iran. His research interests include Mechatronics, Robotics, Social Robotics, Pattern Recognition, Computer Vision, Artificial Intelligence, Machine Learning, and their applications in Human-Robot Interaction.

Ahmad Nourbala graduated with a Mechanical Engineering M.Sc. from the Sharif University of Technology. His research interests are Mechanical Design, Manufacturing, System Dynamics, and Control.

Zeynab Rokhi received her M.Sc. degree in Mechanical Engineering from the Sharif University of Technology. Her research interests are Social Robotics, Computer Vision, Deep learning, Machine learning, and Human-Robot Interaction.

Ali F. Meghdari is a Professor Emeritus of Mechanical Engineering and Robotics at Sharif University of Technology (SUT) in Tehran. Professor Meghdari has performed extensive research in various areas of robotics; social and cognitive robotics, mechatronics, bio-robotics, and modeling of biomechanical systems. He has been the recipient of various scholarships and awards, the latest being: the 2012 Allameh Tabatabaei distinguished professorship award by the National Elites Foundation of Iran (BMN), the 2001 Mechanical Engineering Distinguished Professorship Award from the Ministry of Science, Research & Technology (MSRT) in Iran, and the 1997 ISESCO Award in Technology from Morocco. He is the founder of the Centre ofExcellence in Design, Robotics, and Automation (CEDRA), an affliate member of the Iranian Academy of Sciences (IAS), a Fellow of the American Society of Mechanical Engineers (ASME), and the Founder and Chancellor of Islamic Azad University- Fereshtegaan International Branch (for students with special needs; primarily the Deaf).

Alireza Taheri is an Assistant Professor of Mechanical Engineering with an emphasis on Social and Cognitive Robotics at Sharif University of Technology, Tehran, Iran. He is the Head of the Social and Cognitive Robotics Lab at Sharif University of Technology.

Minoo Alemi received her Ph.D. in Applied Linguistics from Allameh Tabataba'i University in 2011. She is currently an Associate Professor and Division Head of Applied Linguistics at the Islamic Azad University, West-Tehran Branch. She is the cofounder of Social Roboticsin Iran, a title she achieved as a Post-Doctoral research associate at the Social Robotics Laboratory of the Sharif University of Technology. Her areas of interest include discourse analysis, interlanguage pragmatics, materials development, and RALL. Dr. Alemi has been the recipient of various teaching and research awards from Sharif University of Technology, Allameh Tabataba'i University, Islamic Azad University, and Int. Conf. on Social Robotics (ICSR-2014).