The Role of Emotive Feedback in Enhancing Learning through Human-Robot Interaction

ANDRÉ SANTOS, FEUP, Portugal
SÉRGIO ESTÊVÃO, FEUP, Portugal

This study explores the role of emotive feedback in enhancing learning through Human-Robot Interaction (HRI). To achieve this, the research introduces Harpo, a tool designed to facilitate the learning of sign language by incorporating emotive feedback into a software application that allows the users to practice and train on the Sign Language Alphabet characters. By incorporating this feedback, the users are asserted on the experience differences in their engagement and their ability to learn the language. The study employs a Convolutional Neural Network (CNN) to detect and classify sign language gestures, in the form of a VGG16 network followed by a Dense-Layer. The results indicate positive feedback from users regarding the overall experience, with particular emphasis on the engagement-enhancing aspects of emotive feedback. Despite challenges with the model and the difficulty in measuring learning outcomes, users found the tool useful, demonstrating the potential for its appliance in teaching sign language. The study discusses the importance of emotive feedback modalities, including gestures, audio, and visual cues, in shaping user perception and engagement. Ethical considerations and the evolving landscape of Human-Robot Interaction in education are also explored, pointing towards future research directions in the field.

CCS Concepts: • **Human-centered computing**; • **Human-robot interaction**; • **Emotive feedback**; • **Learning enhancement**; • **Educational technology**; • **Technology-enhanced learning**; • **Emotion recognition in robots**; • **User engagement in HRI**;

ACM Reference Format:
André Santos and Sérgio Estêvão. 2023. The Role of Emotive Feedback in Enhancing Learning through Human-Robot Interaction. 1, 1 (January 2023), 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Human-robot interaction (HRI) is a broad research field dedicated to studying and designing the usage of robot systems alongside humans on some recurrent tasks and the evaluation of their viability. In order to fully comprehend all the specifics of this area, one must intersect various domains, ranging from the technical aspects like robotics or artificial intelligence to less computer-related problems like ethics and psychology. Over the years, robots have been increasingly joined with humans to perform different tasks, along with the evolution of their capacities and abilities, however, the main focus of this area is their inclusion in society as social robots, entities capable of engaging, communicating and being introducing usefulness in areas as learning or psychological health care [2].

Despite the existence of some obstacles, the introduction of educational robots into learning programs has been a big topic in society. These can be useful in promoting active engagement, problem-solving, and collaboration among students. They are usually perceived as a social enhancing tool, as interacting with robots can be less intimidating than interacting with peers. Besides, some skills as creativity and critical thinking are not discarded or less worked than the regular environments[4]. Some applications have tried to employ these interactive components in a learning environment, in order to improve user learning. A known example of this is the Duolingo app to learn languages [3]. By creating a figure that interacts with the user with audio, text, and movements, the student is perceived to be more committed and engaged to keep using the application and effectively learning. On a more robot-oriented project, L2TOR[12], SoftBank's NAO robot was used as a tutor supporting preschool children in learning a second language. The results were positive and the children included in the study were generally able to learn some words in a new language.

While there regular language learning has received attention in recent years, the same cannot really be said about sign language. This is the primary language of communication for groups of people with disabilities, however, the study of sign language remains underrepresented in mainstream education and its growth has been stagnating. This is a big obstacle and deaf individuals are generally excluded from learning opportunities. Sign language also has the potential to be used in various scenarios as scuba diving, communicating in quiet areas, or international dialogue.

Despite the current referred issues, HRI can be seen as an opportunity to foment the learning of Sign Language. Even though learning can be seen as not so entertaining, the inclusion of an interactive environment, similar to other already mentioned tools, could be the way to create a dynamic environment and break down barriers for people to learn and therefore facilitate the access of education for deaf individuals. As such, Harpo is introduced as a solution to all these issues. This study will investigate how the presence of an agent in a sign learning tool will affect the cognitive and physical learning process for a set of users.

2 BACKGROUND

In order to provide some background on the importance of emotional feedback in learning, on the different approaches by which this can be communicated to a human, and on the sign language classifier model works, this section will review a few information from different sources that will create a base for the exploratory study.

2.1 Emotive Feedback in HRI

Definition and Importance: Define emotive feedback in the context of HRI. Discuss why emotional responses from an avatar or

Authors' addresses: André Santos, FEUP, Porto, Portugal, EMAIL; Sérgio Estêvão, FEUP, Porto, Portugal, smestevao11@gmail.com.

XXXX-XXXX/2023/1-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

robot can be significant in a learning environment. Mechanisms of Emotive Feedback: Explore how emotive feedback is conveyed through virtual avatars or robots – facial expressions, gestures, voice modulations, etc.

Emotive feedback in the context of Human-Robot Interaction (HRI) refers to the ability of a robot to express and convey emotions in response to the emotional cues it perceives from humans or the environment. The robot provides feedback that complements the emotional state of the user, creating a more emotionally intelligent relationship between both. Humans are inherently social creatures, driven by a fundamental need for connection, interaction, and the establishment of meaningful relationships with others. The employment of emotion in an interaction between a person and a computer stimulates the establishment of a more trusty connection. Emotive feedback can be conveyed through various modalities, including facial expressions, vocal intonation, body language, and other non-verbal cues.

- **Gestures**: Are fundamental in non-verbal communication. It can be a strong influence on how a message is perceived. Usually, it is a big compliment on the clarity and meaning of communication. It could be seen as an extra form of engagement by a robot.
- **Audio**: Sound cues are an important form of communication since they are the only trigger that makes use of the hearing sense. It can be really important in providing feedback for an action, whether it be more informative or more emotive. They are also really relevant in providing context on how a message is perceived. In the case of robotics, they are a big aid in establishing communication and interaction with humans since they influence how much a person perceives a robot as human-like.
- **Lightning / Colors**: Simple form of communication that takes advantage of what the humans perceive as positive or negative, bright or dark, etc. It is a simple form of feeding approval and changing the environmental perception of the user. For example, green is perceived as success and may induce a greater emotion of accomplishment.
- **Touch**: Haptic feedback provides tactile sensations. When introduced in a robot, it can have negative effects since this type of sensation can be perceived as too unfamiliar for humans since it is already between two persons, let alone a robot. However, there are many forms of providing haptic feedback, such as vibrations in a controller or the robot itself.
- **Emotion**: Also mainly on the visual feedback side, but one of the best ways of engaging with a human. Generally, people are really sensitive to facial expressions, smiles, eye movements, etc. While it can also be perceived as weird in a robot, emotive communication is one of the biggest wall breakers in the human-robot interaction since it changes the way the human perceives the robot as a machine. It can also be used with other senses.

Non-verbal communication is a very important part of the way humans communicate. In order to reduce the distance that still separates the human-robot interactions, the expression of emotion is fundamental.[9]

Despite this, there is a need for care in the way these emotions are employed in robots. As referred in [8], the human likeness for an entity is not linear, depending on its human resemblance. The function known as Uncanny Valley shows that there's a point at which a robot becoming more similar to a human will drop its positive perception until it is similar. Because of that, when employing these interactive cues, especially on an entity such as a robot, it is important to measure how the emotions are being communicated and if the humans perceive them positively.

2.2 Impact on Learning and Engagement

When it comes to learning, it is also pretty clear that a strong expression of emotions is a recipe for more engagement from the "student" point of view and, therefore, a higher learning rate. With the introduction of computers, electronics and robots into the current society, in the form of different learning methods and, for example, learning from home, the form teaching is performed needs to be adjusted to match the learner's needs. In [5], a study was conducted to evaluate how students would behave with different types of feedback. "Results showed that emotional feedback decreased confusion, triggered intrinsic motivation, and enhanced agent perception". The attention span and motivation-related feelings are then proven to be improved despite the cognitive load-related sensations not being reduced. In [11], another study was conducted on the emotions shown by students depending on the feedback. The work mentions the thin line between the management of emotion depending on the context and that feedback can work both ways. However, the general conclusion is that this feedback is positive and can also be seen as important to help people manage their emotions.

2.3 CNN

A Convolutional Neural Network (CNN) is a class of deep neural networks most commonly applied to image analysis. This network uses a mathematical technique called Convolution. It is generally an operation between two functions that produce a third that expresses how the shape of one is modified by the other.

Generally, the goal of a CNN is to reduce images into a form that makes it easier to extract features and similarities, which allows the processing of images without losing details, whether they be bigger or smaller.

In every layer, a filter/kernel of a specific size is applied to the image, a group of pixels, and the convolution operations are executed, taking the information on those pixels and forwarding them to the next layer. This operation is often performed in different layers, building the Neural Network.

The first layer usually extracts basic features such as horizontal or diagonal edges. This output is passed on to the next layer, which detects more complex features such as corners. As the layers get deeper, more complex features, such as faces, eyes, etc, can be detected.

Based on the activation map of the final convolution layer, the classification layer outputs a set of confidence scores (values between 0 and 1) that define how likely the image is to belong to a class. The highest value between the considered classes is the network prediction, but similar higher values can be accounted for as well, as the CNN is uncertain[7].

Many different types of networks belong to this category and are used for different approaches and with also different results1.

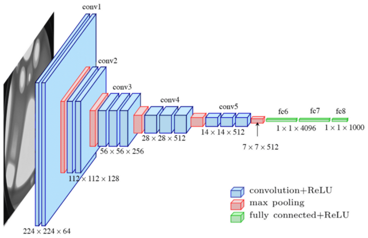

Fig. 1. VGG16 Architecture[6]

3 HARPO - SIGN LANGUAGE CASE STUDY

To perform the study on the effect of emotive feedback in a sign language learning environment, the software is to be designed to provide a backbone for the tests. This application can be divided into two main areas: the sign language detector and the interface for the user to learn and experiment.

3.1 Application Overview

Application Overview: Provide a detailed description of your application, emphasizing the interactive and emotive aspects of the avatar or character used.

3.2 Sign Language Detector

Detecting and classifying sign language gestures prompted by a user in a camera sensor involves different approaches and techniques. The general approach in these tasks is to use deep learning to classify the image as a certain letter(class). To do so, a Convolutional Neural Network(CNN) is to be employed to learn hierarchical features from the images automatically. CNNs are particularly effective in capturing spatial patterns and values, making them well-suited for image classification tasks. Usually, the best approach is to use a pre-trained model, whose results are usually better than manually defined networks. Examples of these are ResNet, VGG16, or DenseNet.

The chosen model was a VGG16 with a Dense Layer with a softmax activation function for multiclass classification of all the possible characters based on the model used in [10]. When evaluating the test dataset, the model converged at around 10 epochs and achieved an average of 91% of precision, recall, and accuracy2.

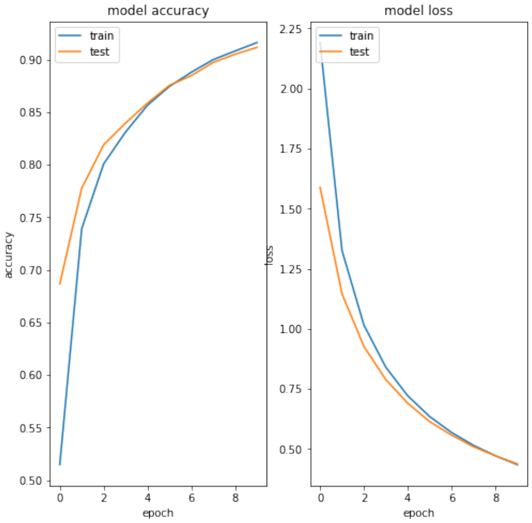

Fig. 2. Model Training Review[10]

The parameters used to train the model were:

- 80/20 Train/Test Split
- Optimizer: Adam
- Learning Rate: 1e4
- Number of Epochs: 10
- Loss function: Categorical cross-entropy
- Main Metric: Accuracy

The dataset was one of the main issues when training the model. There are different sets of data around the internet, but usually, the main problem is that they are taken from video sequences, which means all the frames have the same light conditions and hand patterns. This is an issue since the model is then trained for similar data sets, and it overfits too much to its data, failing when prompted with new inputs. This could be tried to be solved by using different datasets or generating their own data, but those are not fault-proof options. The first training sessions always ended with this outcome. Good performance for the model data, poor when classifying with different images. Sometimes, the letter was not even similar. This was solved by using the referred model [10] with the dataset [1]. It contains 87000 images for 29 different elements, 26 letters + the ’space’, ’delete’, and ’nothing’. Despite not being a clearly different dataset when compared to the others, it is a better solution, which resulted in a model with understandable and valid results.

Even though the model achieves good results on a 29-multi-classification task, it is easy for the model to be unsure which sign it is classifying, especially when distinguishing between hand signs. Some of the letters in the alphabet are similar and easy to be mistaken for, even by humans. The ’J’ and ’Z’ letters are extra hard to predict since they require movement. When considering issues such as camera quality, light conditions, and lousy backgrounds, the task becomes difficult, no matter how good the model is.

To solve this, when an image is prompted, if the activation for some class is above a certain threshold, it is accepted, even if it was not the highest rated probability and, therefore, the model prediction.

This way, the perception and sensing problems can be contoured by using a not-too-restrictive approach.

3.3 User Interaction Mode

User Interaction Model: Explain how users interact with the application and how the system provides emotive feedback based on user performance.

The User Interaction Model is centered around the user's interaction with the intuitive and responsive interface that allows users to engage seamlessly with the emotion recognition system. This interface revolves around submitting real-time images captured to an inference model that performs emotion analysis.

Technically speaking, the application is divided into two segments. The backend, which receives the images periodically and can infer if an image contains a gesture for a specific hand sign based on whether the prediction value for the current training letter fulfills a threshold, as well as attributes an emotion to the prediction based on how far away the image if from the threshold. The emotions can range from sad, neutral, happy, and excited. Being sad when the user is the furthest from achieving the correct sign language gesture, and excited when the user can perform the hand gesture and fulfill the minimum threshold successfully.

On the other hand, the frontend, the part of the application that the user directly interacts with, is composed of a camera showcase, an emotional feedback element on the top-left, an image depicting the correct hand gesture for the chosen letter on the top-right, and an input box to choose the desired letter to learn in sign language on the bottom. The frontend every second sends an image to the backend, updating the emotional feedback element, which in turn gives the user information on how his performance is. With the input box, the user can change which letter he tries to learn at any time, and the hand gesture depiction on the top right updates accordingly.

In figure 3, it is possible to access the application architecture and technologies used, namely React,js for the frontend, Python with Flask for the backend, and Tensorflow for the training of the AI model.

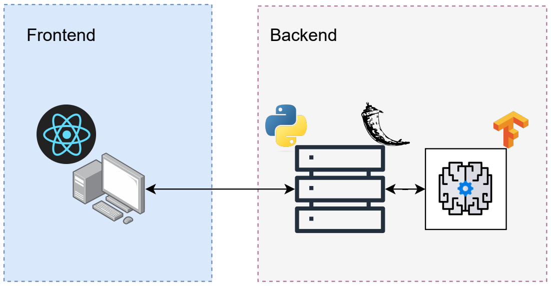

Fig. 3. Application Architecture

4 RESULTS

Considering the mentioned constraints, the model is able to assert whether a sign is similar to a pretended one. By doing so, the users are able to try to compose the right character and learn as they repeat this process. The main goal of this study was to evaluate how the interaction with Harpo changed their learning behavior. Since the alphabet still has a large amount of characters and users are not familiar with the language, hypothesizing on their actual learning speed was hard, since this would take a little time. Therefore, the focus ended up being the user perception of the effect of the emotive feedback on their performance. A few evaluation metrics and questions were created to try to get a grasp of the user experience, including emoji prompts, audio feedback, the presence of human-like figures, and the appearance of cartoons. In total, 6 university students took part in the study, all ranging from 18 to 22, half masculine and half feminine.

4.1 Qualitative Evaluation

As expected, when asking the users for an opinion on how they felt by introducing the feedback, the responses were not too relevant or useful. The general comment would fall down on the "funny" family of adjectives and the users would feel smiley and happy with the interaction, but it was hard to assert an actual improvement in the user learning.

4.2 Quantitative Evaluation

In order to have a quantitative evaluation that could be used to get more exact results, the users responded to the different queries with a value from 1 to 5, where 1 is "I hated it" and 5 is "I loved it". The questions were:

- Q1: How did you find your experience? (No Feedback)
- Q2: How did the emoji prompt affect your experience?
- Q3: How did the audio feedback affect your experience?
- Q4: How did the presence of a human-like figure affect your experience?
- Q5: How did the appearance of cartoons affect your experience?
- Q6: How easy was learning sign language? (No Feedback)
- Q7: How easy was learning sign language? (Generally w/ Feedback)

Questions	**Subjects**					
	A	**B**	**C**	**D**	**E**	**F**
Q1	4	4	3	4	5	3
Q2	3	4	3	3	3	3
Q3	2	4	3	4	4	3
Q4	3	4	3	3	4	3
Q5	3	4	3	3	3	3
Q6	1	3	2	3	2	2
Q7	1	3	2	3	3	2

Table 1. Survey on Subjects Perception of the Application

In general, people stayed pretty neutral, within the 2 and 4 values, which was expected. When asked on whether the general experience, the results were positive, since the overall opinion was of enjoyment, as stated in the previous subsection.

When asked about the different feedback effects on the user experience, the users varied most in audio feedback, which is a more different type than the rest, and as such more polarizing. The visual cues, all stayed pretty neutral. However, overall people were positive about the usage of emotive cues within the application.

Finally, when asked whether the feedback affected their learning experience, the answers were almost the same between the no feedback and the feedback alternatives. This is understandable as it is hard to perceive whether this emotion transmission has an effect on learning when comparing it on the spot, especially when there is still a big learning curve. The overall opinions were a bit shifted to the hardness of the task, usually because of the big learning curve for that many characters.

Because of the model uncertainty, and the overall difficulty in succeeding in the sign language evaluation, a 'Wizard of Oz' experiment was also carried out, by altering the actual response to a streak of successes. While this didn't measure any effect on the learning nor on the emotional feedback, and the evaluation of this being qualitative, since the users were not allowed to know of this variation, their opinion on how difficult learning was changed after the increase in the success rate.

5 DISCUSSION

Despite the sign language evaluation already being a not so simple task, it is still possible to make assertions on how having emotive feedback affected the user learning and their perception of it. The model volatility was an issue and impeded the responsiveness of the program a few times, which caused some disbelief in the tool. The low amount of participation in the survey could have impacted the results, but the general idea should stay within the same parameters, given the existing conditions. In spite of this, the results were still positive, and conclusions could be taken from the introduction of this feedback, even if not on a large scale.

5.1 User Feedback

While there can be many ways of understanding how a human is reacting to an interaction, it is not a trivial task. One of the issues is related to how each person shows their feelings, which varies a lot and can be misunderstood. In order to address this, an evaluation survey can be performed, asking the user to explain or rate the interaction, comparing the two alternatives, and making or validating statements about it. However, this creates another issue called bias. When evaluating a task like this, the first time perception is really important, since the users are acting on a novel tool, and their response to feedback differs because of that. It is hard to compare the usage of different feedback since the user will get used to the platform and thus not be able to really understand the authentic emotions he developed. Despite this, the study still tried to evaluate these metrics by inquiring on concrete opinions and ratings.

5.2 Robot or Application

It may seem clear whether a user is experimenting with a robot or just an interactive application, however, the line that separates the two types of feedback is not so big. Focusing more on the software component will bring positive aspects such as accessibility and less strangeness for the study. However, the feedback the software can provide is always limited by the hardware. A computer using a camera sensor will be limited to visual feedback. If a speaker is available, audio feedback can be introduced. If a robot with physical movements is utilized, haptic feedback can also be on the table. However, the deeper the desire is to introduce reality in these interactions, the harder it is to have success, given the robotic limitations we have today. It may be difficult to test these metrics and evaluate the user interaction with them, but while it is not developed to its full extent, it is possible to clearly visualize trends in studies like this. AI development will enhance the ability to incorporate emotional responses based on user behavior with higher quality, while hardware upgrades will also turn robots more similar and human-like.

5.3 Ethical Considerations

While technology has somewhat swiftly been incorporated into educational settings, and its usage has been proven to be beneficial in some aspects, there is still a long way for human-robot interaction to evolve, in order for robots or machines to teach with the same emotional background as a regular teacher. There will also be an ethical and moral questioning barrier on the usage and trade-offs of incorporating robots into our daily tasks, specifically in these tasks where the good development of social skills is required. Besides, user dependency and privacy concerns are really present these days and these issues will generate a big discussion and barrier to this advancement.

6 CONCLUSION

In summary, the task of learning sign language by providing emotive HRI was complex, but revealing. The model's difficulty in clearly identifying a specific class introduced a barrier to the responsiveness and assertiveness of the classification. Besides, the difficulty in creating meaningful feedback that would influence user performance, as long as measuring it was also a big obstacle.

Despite this, the general opinion was that it was a useful tool, and could be seen as a way of teaching sign language, especially with feedback, since it enhances user engagement, despite no learning differences could be asserted. As we gaze into the future, the employment of emotive feedback in HRI seems to be effective and a trend to maintain, especially in educational environments, in spite of the existence of some barriers. With the development of AI and hardware enhancements, the scenario can change and the investigation on the area should be maintained, in order to redefine the relation humans and robots have these days.

REFERENCES

[1] [n. d.]. ASL Alphabet. https://doi.org/10.34740/KAGGLE/DSV/29550

[2] Kerstin Dautenhahn and Aude Billard. 1999. Bringing up robots or—the psychology of socially intelligent robots: from theory to implementation. In *Proceedings of the Third Annual Conference on Autonomous Agents* (Seattle, Washington, USA) *(AGENTS '99)*. Association for Computing Machinery, New York, NY, USA, 366–367. https://doi.org/10.1145/301136.301237

[3] Duolingo Inc. [n. d.]. *Duolingo*. https://www.duolingo.com/

[4] ITU AI for Good Global Summit. [n. d.]. *The Future of Educational Robotics: Enhancing Education, Bridging the Digital Divide, and Supporting Diverse Learners*. https://aiforgood.itu.int/the-future-of-educational-robotics-enhancing-education-bridging-the-digital-divide-and-supporting-diverse-learners/

[5] Yueru Lang, Ke Xie, Shaoying Gong, Yanqing Wang, and Yang Cao. 2022. The Impact of Emotional Feedback and Elaborated Feedback of a Pedagogical Agent

on Multimedia Learning. *Frontiers in Psychology* 13 (2022). https://doi.org/10.3389/fpsyg.2022.810194

[6] Khuyen Le. 2021. *An overview of VGG16 and NiN models.* https://medium.com/mlearning-ai/an-overview-of-vgg16-and-nin-models-96e4bf398484

[7] Manav Mandal. 2023. *Introduction to Convolutional Neural Networks (CNN).* https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/

[8] Masahiro Mori. 2012. Uncanny Valley. https://spectrum.ieee.org/the-uncanny-valley

[9] Sharon Oviatt, Björn Schuller, Philip R. Cohen, Daniel Sonntag, Gerasimos Potamianos, and Antonio Krüger (Eds.). 2017. *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 1.* Vol. 14. Association for Computing Machinery and Morgan & Claypool.

[10] Paul Timothy Mooney. 2018. *Interpret Sign Language with Deep Learning.*

[11] Anna Rowe, Julie Fitness, and Leigh Wood. 2014. The role and functionality of emotions in feedback at university: A qualitative study. *Australian Educational Researcher* 41 (07 2014), 283–309. https://doi.org/10.1007/s13384-013-0135-7

[12] Paul Vogt, Rianne van den Berghe, Mirjam de Haas, Laura Hoffmann, Junko Kanero, Ezgi Mamus, Jean-Marc Montanier, Cansu Oranç, Ora Oudgenoeg-Paz, Daniel H. Garcia, Fotios Papadopoulos, Thorsten Schodde, Josje Verhagen, Chris D. Wallbridge, Bram Willemsen, Jan de Wit, Tony Belpaeme, Tilbe Göksun, Stefan Kopp, Emiel Krahmer, Aylin C. Küntay, Paul Leseman, and Amit K. Pandey. 2019. Second Language Tutoring using Social Robots. A Large-Scale Study. In *Proceedings of the 2019 ACM/IEEE International Conference on Human-Robot Interaction (HRI 2019).*

Received January 23, 2024