Increasing Telepresence Robot Operator Awareness of Speaking Volume Appropriateness: Initial Model Development

Matthew Rueben
Thomas Groechel
Yulun Zhang
Maja J. Matarić
Department of Computer Science
University of Southern California
Los Angeles, California, USA
{mrueben,groechel,yulunzha,mataric}@usc.edu

Gisele Ragusa
Division of Engineering Education
Viterbi School of Engineering
University of Southern California
Los Angeles, California, USA
ragusa@usc.edu

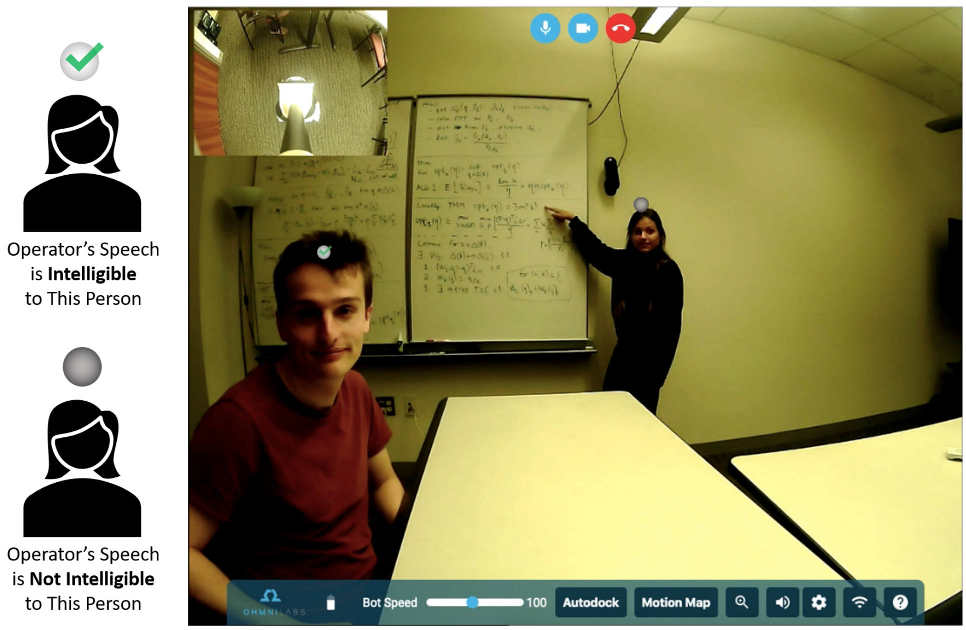

Figure 1: Screenshot of the Ohmni telepresence robot operator interface (right) and cartoon of the elements added by our system (left; not shown to robot operators). Our system displays one of two icons over each detected face to indicate whether that listener can understand what the operator is saying at the current volume: a green check mark inside a white circle (closer person, sitting at left) or an empty gray circle (farther person, standing at whiteboard).

ABSTRACT

Telepresence robots could help homebound students to be physically embodied and socially connected in the classroom. However, most telepresence robots do not provide their operators with information about whether their speaking volume is appropriate in the remote context. We are investigating how operators could benefit from live feedback about speaking volume appropriateness as part of our ongoing research on using remote presence robots to improve education and social connectedness for students experiencing extended absences from school. This preliminary report describes (1) the development of a model of speaking volume appropriateness to provide this feedback, (2) implementation of a feedback element in the operator's user interface, and (3) plans for long-term deployment to assess impacts on the social and educational experience of homebound high school students.

HRI '20 Companion, March 23–26, 2020, Cambridge, United Kingdom

ACM ISBN 978-1-4503-7057-8/20/03.
https://doi.org/10.1145/3371382.3378388

ACM Reference Format:
Matthew Rueben, Thomas Groechel, Yulun Zhang, Maja J. Matarić, and Gisele Ragusa. 2020. Increasing Telepresence Robot Operator Awareness of Speaking Volume Appropriateness: Initial Model Development. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20 Companion), March 23–26, 2020, Cambridge, United Kingdom.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3371382.3378388

1 INTRODUCTION

Interacting with peers and teachers in the classroom is essential for the educational and social development of K-12 students; unfortunately, more than a quarter of children in the United States miss significant amounts of school each year [4]. Causes for extended absences include chronic illnesses, mental health challenges, and behavioral issues. Interventions such as homeschooling, individualized tutoring, and online learning cannot fully replace the social and educational value of learning alongside one's peers [1].

Telepresence robots have the potential to give absent students more direct access the K-12 classroom, including interactions with peers and teachers over a video call as well as a physically present, mobile surrogate body (the robot) that the remote student can drive. To exploit this opportunity, however, the telepresence robots that are commercially available today require some modification for classroom use [2]. One difficulty for the robot operator is discerning whether their speaking volume is appropriate for the remote context. The telepresence robot operator is often unaware of the volume at which their voice is being projected from the robot's speakers, perhaps because it is not played back to the operator with adequate volume or fidelity [5]. Previous work has modelled whether a robot will be able to understand other people's speech, but not the reverse [3]. *This late-breaking report describes the development of a model of speaking volume appropriateness to investigate whether live feedback improves the classroom experience for robot operators.*

2 DESIGN GOALS AND APPROACH

The goal of this work is to investigate how homebound telepresence robot operators could benefit from more feedback about whether their speech is loud enough to be understood by people in the classroom. We do not know, however, to whom the operator is trying to speak—e.g., to the teacher across the room or to the student right next to the robot. Our system will therefore give feedback about speaking volume appropriateness for *every possible listener* that the robot can detect.

Furthermore, we found through testing that the robot's speaker cannot transmit an operator's voice loudly enough to be uncomfortable to a listener, even if the listener is directly in front of the robot. Our interface element will therefore display only whether the speech is too quiet for each listener to understand.

3 IMPLEMENTATION FOR DEPLOYMENT

An interface element controlled by our model of speaking volume appropriateness was implemented for the upcoming classroom deployment. The element overlays feedback about speaking volume appropriateness on the web operator interface for the Ohmni telepresence robot used for this project. As shown in Figure 1, an icon is displayed over each detected face in the robot's field of view. A semi-transparent gray circle indicates that the operator's speech is not intelligible to that listener (or the operator is not talking); the icon changes into a green check mark when the speech becomes loud enough to be understood by that listener.

We developed a computational model of speech volume appropriateness in order to determine which icon to display. First, the raw speech volume at the operator's microphone and the robot speaker volume setting that the operator can adjust are used to calculate the volume of the operator's voice as it leaves the robot's speaker. That volume level is then compared to the minimum intelligible volume level produced by the model for each potential listener using two additional inputs: the distance to each potential listener (calculated from the detected height of the potential listener's face) and the ambient noise level in the room (measured by the robot's microphone).

4 TRAINING DATA COLLECTION AND MODEL FITTING

Data were collected from several participants from our lab to characterize the relationship between the distance to the listener, ambient noise level, and minimum intelligible volume. The distance to listener and ambient noise volumes were manipulated experimentally. A pre-recorded spoken phrase was played through the robot's speaker and pilot participants were instructed to adjust the robot's volume setting until they could just barely understand what was being said. The mean volume of this sound file was computed and then used along with the chosen robot volume setting to calculate the minimum intelligible volume for that pilot participant at the current distance and ambient noise level.

The robot volume settings chosen by the pilot participants describe the boundary between the two volume bins: unintelligible and intelligible. This boundary was found to be roughly planar when the speaker volume is in decibels, the ambient noise level is in the native units of the robot's microphone, and the natural logarithm of the distance in meters is used. The minimum intelligible volume is in decibels:

$$\begin{aligned} MinIntelligibleVolume = 0.01168 * AmbientVolume + \\ 6.90635 * ln(Distance) + 49.40575 \end{aligned} \tag{1}$$

5 PLANNED EVALUATION IN LONG-TERM DEPLOYMENTS

The effects of the feedback provided by our system will be measured over the course of several long-term deployments helping homebound high school students attend class remotely. Questionnaire items administered to each participant will measure their awareness of whether their voice volume was socially appropriate in the classroom. Also, video recordings of the interactions will be annotated for instances of the teacher and classmates asking the operator to adjust the volume, as well as for times when the operator adjusts it on their own. Findings from these evaluations are expected to inform the design of speaking volume feedback on commercially available telepresence robots and help build an understanding of how to create better social and educational experiences for students experiencing extended absences from school.

ACKNOWLEDGMENTS

This work was supported by an NSF NRI grant for "Socially Aware, Expressive, and Personalized Mobile Remote Presence: Co-Robots as Gateways to Access to K-12 In-School Education", NSF IIS-1528121. We also acknowledge OhmniLabs for providing the Ohmni robot platform for this work as well as excellent technical support.

REFERENCES

[1] John Bransford, Reed Stevens, Dan Schwartz, Andy Meltzoff, Roy Pea, Jeremy Roschelle, Nancy Vye, Pat Kuhl, Philip Bell, Brigid Barron, Byron Reeves, and Nora Sabelli. 2006. Learning theories and education: Toward a decade of synergy. In *Handbook of Educational Psychology* (2nd ed.), P. Alexander and P. Winne (Eds.). Lawrence Erlbaum Associates Publishers, Mahwah, NJ.

[2] Elizabeth Cha, Qandeel Sajid, and Maja Matarić. 2016. Enabling access to K-12 education with mobile remote presence. In *2016 AAAI Spring Symposium Series*.

[3] Ross Mead and Maja J Mataric. 2014. Probabilistic models of proxemics for spatially situated communication in HRI. In *Int. Conf. on Human-Robot Interaction, Algorithmic Human-Robot Interaction Workshop.*

[4] U.S. Department of Health and Human Services. 2012. The 2011–12 national survey of children's health. Retrieved December 1, 2017 from http://www.childhealthdata.org/learn/NSCH.

[5] Andreas Paepcke, Bianca Soto, Leila Takayama, Frank Koenig, and Blaise Gassend. 2011. Yelling in the hall: using sidetone to address a problem with mobile remote presence systems. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 107–116.