Comparing the Effect of Different Styles of Voice on Children's Engagement With a Virtual Robot: A Preliminary Study

Romain Vallée & Lucas Prégaldiny

ACHI 2023

Research Context

Lucas Prégaldiny's MSc thesis

Romain Vallée's PhD thesis

Véronique Aubergé's research

+ Émilie Cénac, Serge Tisseron, Olivier Aycard

Enchanted Tools

cité sciences et industrie

Lutin Userlab
Cité des sciences et de l'industrie

Research Objectives

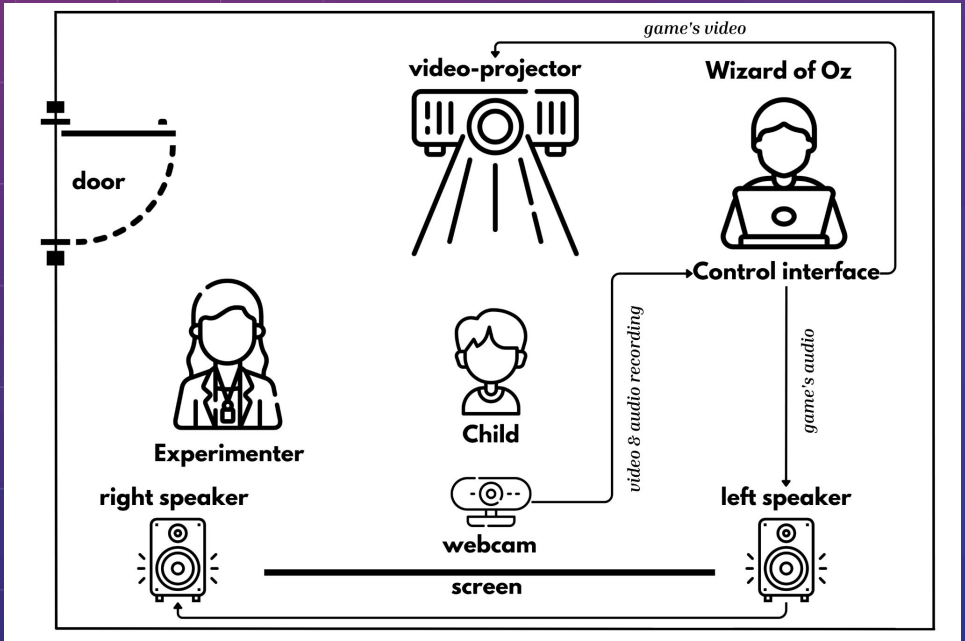
Research hypothesis:

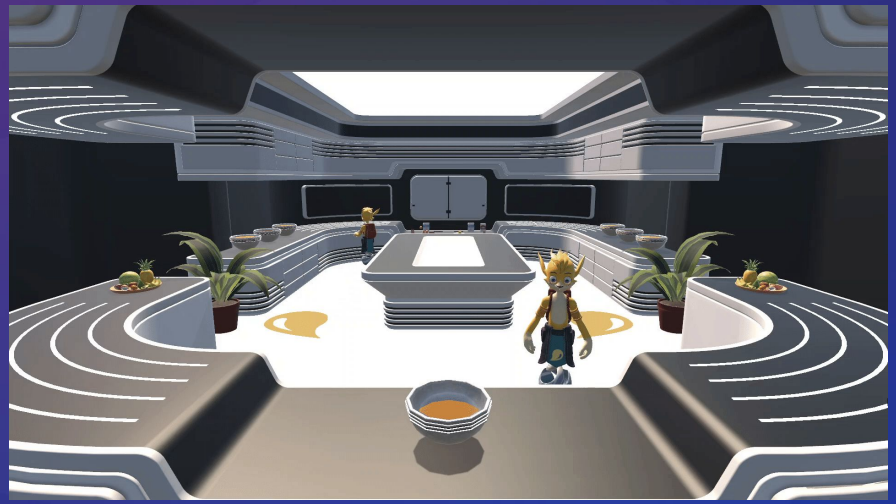
In a playful cooking task involving children interacting with virtual robots, a robot using an expressive voice and non-lexical speech elements will elicit more engagement, trust or interest than a robot using only lexical speech elements and a constant prosodic modality.

Voice A	Colloquial enunciation, aiming for a playful, dynamic style, exaggeratedly child-like cartoon voice: modal or tense voice (tensed), fast-paced, high pitched (mean fundamental frequency F0 = 320 Hz)
Voice B	Same instructions and prosodic values as voice A, but including non-lexical socio-affective vocal primitives consistent with the global prosody (vocal bursts, grunts, onomatopoeia, etc.) (mean F0 = 320 Hz)
Voice C	An acted voice simulating “stereotypical” synthetic voices (e.g., Alexa), i.e., globally breathy without attitude variations and without non-lexical vocal elements: systematically breathy voice, slow rhythm, lower pitch (mean F0=250 Hz)

The audio files used in the experiment are available at https://lpy-et.github.io/ACHI2023/

Experimental set-up





Link to a video of the game

Game features

- 6 ingredients for a chocolate cake
- 4 minutes of average duration

Questions to children

Before the game

- Have you ever played video games?
- Have you ever cooked at home?
- Have you ever seen robots?
- What are robots used for?

After the game – general questions

- Did you like this game?
- Was it good that the robots helped you with this recipe?
- Was it good that I helped you?
- Is there a robot who helped you more?
- Is there a robot you liked more?

Questions to children

After the game – more specific questions

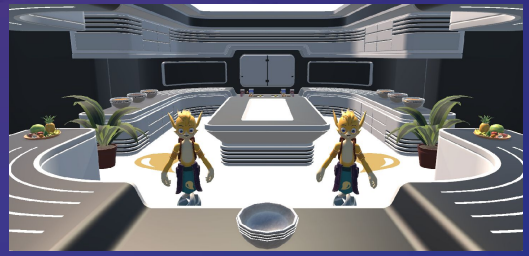
- Do these robots talk?
- Do they talk the same way?
- How do they talk?

- Is there a robot you understood better?
- Is there a robot you preferred to talk to?

- Would you lend a precious object or a toy to one of the robots?
- Would you let one of the robots enter your bedroom?

Final choice

- “Take the sticker which represents the robot you liked best”



Stickers

Participants

30 children from 6 to 10 years old

- 35% girls / 65% boys
- average age: 7.8 years (± 1.7 SD)

3 groups of 10 children

- A/B group interacted only with voices A & B
- B/C group only with voices B & C
- A/C group only with voices A & C

Children’s involvement

- 100% said they enjoyed the game
- 95% said they appreciated the robots’ help
- 75% said they appreciated the experimenter’s help

Perceived differences in voices

85% of the children noticed that the two virtual robots did not speak in the same way

Difficult to identify children's preferences from these data

TABLE I
QUALIFIERS FOR EACH VOICE BY NUMBER OF OCCURRENCES (NB)

Voice A		Voice B		Voice C	
qualifier	**nb**	**qualifier**	**nb**	**qualifier**	**nb**
higher than C	3	higher than C	3	lower than B	3
lower than B	2	higher than A	2	lower than A	2
higher than B	2	softer than C	1	male	2
female	2	nicer than C	1	bigger than B	1
high	1	smaller than C	1	slow	1
fast	1	better than C	1	alien	1
enthusiastic	1	lower than A	1	not too fast nor too slow	1
		laughing	1		
		shouting	1		
		unpleasant	1		
		cheerful	1		
		female	1		

reading hint: voice A is reported as higher than voice C by 3 children

Engagement/Trust/Preference - Score calculation

Method

- Assign X/10 points If a robot is chosen or preferred by X%
- Example for voice A in group A/C:
  - 1 point (Q1 : 10% for only A)
  - 2 points (Q2 : 20% for only A)
  - 2 points (Q3 : 20% for only A)
  - 4 points (Q5 : 40% for both A and C)
  - 1 point (Q5 : 10% for only A)
  - 1 point (Q6 :10% for only A)
  - 6 points (Q6 : 60% for both A and C)
  - 7 points (Sticker : 70% for A)
  - Total of 24 points

group	answer	Q1	Q2	Q3	Q4	Q5	Q6	sticker	score
A/B	none	90%	70%	90%	70%	10%	30%		
	both					40%	30%		
	A	10%	30%	10%	10%	40%	30%	**60%**	**26**
	B				20%	10%	10%	40%	15
A/C	none	80%	80%	80%	90%	40%	30%		
	both					40%	60%		
	A	10%	20%	20%		10%	10%	**70%**	**24**
	C	10%			10%	10%		30%	16
B/C	none	50%	30%	50%	40%	40%	30%		
	both								
	B	30%	20%	40%	20%	20%	50%	**80%**	**26**
	C	20%	50%	10%	40%	40%	20%	20%	20

Q1 Is there a robot who helped you more?
Q2 Is there a robot you liked more?
Q3 Is there a robot you understood better?
Q4 Is there a robot you preferred to talk to?
Q5 Would you lend a precious object or a toy to one of the robots?
Q6 Would you let one of the robots enter your bedroom?

Effect of non-lexical primitives

- Tendency to prefer expressive voices when opposed to the synthetic one: 50 points for A+B (24 for **A**/C and 26 for **B**/C) *vs* 36 points for C (16 in A/C and 20 in B/C)
- The frequencies of answers “none” to Q1, Q2, Q3 and Q4 change substantially between A/C and B/C, which suggests a dividing effect of the non-lexical elements
- Voice A was always preferred: total score of 50 points, when B had 41 and C had 36

group	answer	Q1	Q2	Q3	Q4	Q5	Q6	sticker	score
A/B	none	90%	70%	90%	70%	10%	30%		
	both					40%	30%		
	A	10%	30%	10%	10%	40%	30%	**60%**	**26**
	B				20%	10%	10%	40%	15
A/C	none	80%	80%	80%	90%	40%	30%		
	both					40%	60%		
	A	10%	20%	20%		10%	10%	**70%**	**24**
	C	10%			10%	10%		30%	16
B/C	none	50%	30%	50%	40%	40%	30%		
	both								
	B	30%	20%	40%	20%	20%	50%	**80%**	**26**
	C	20%	50%	10%	40%	40%	20%	20%	20

Q1 Is there a robot who helped you more?
Q2 Is there a robot you liked more?
Q3 Is there a robot you understood better?
Q4 Is there a robot you preferred to talk to?
Q5 Would you lend a precious object or a toy to one of the robots?
Q6 Would you let one of the robots enter your bedroom?

Relevance of holistic gestural behavior

- The preference of a child towards a virtual robot is not only verbalized but more generally *gesturalized*, when the child gets to choose a physical sticker
- Notable for the B/C group: similar scores of B and C for the question-answering, but the B sticker is chosen 8/10 times

Stickers

Conclusions

- Voice seems to have a significant impact in the context of a video game for children, where many other parameters intervene (the visual aspect of the game, its playability, its novelty, the presence of an adult at the child's side...), which could have strongly limited this impact
- The factors of voices A and B place the child in an expressiveness (e.g., "enthusiastic"), which attract their preference, and don't have the intimacy of voice C (which they sometimes describe as "polite", "friendly")
- Non-lexical speech elements seem to install a relational space different from the one installed by strictly lexical elements
- This will guide further research devoted to analyzing more precisely the relational effects pointed by the inconsistency when comparing directly voice A and B