

MultiVRse Heuristic Evaluation

Prototype Description

MultiVRse helps users learn English as a second language (ESL) by providing VR simulations where users can practice English; for this heuristic evaluation, the VR simulation provided was talking to a mechanic fixing your car in a Toyota dealership.

Task

Schedule your car to be fixed as soon as possible because you need to drive it to work and your warranty ends soon.

Versions Note

Evaluator A saw the dancing mechanic. There was no option to state when you wanted to exit. There was not a paper introduction describing the situation in more detail. This simulation did not ask for the phone number.

Evaluator B saw the zombie (with tattered clothing, hunched back, and no shoes) and dancing mechanic. There was no option to state when you wanted to exit. There was a paper introduction describing the situation in more detail. This simulation asked for the phone number.

Evaluators C & D saw a version of the simulation that featured no animations. The mechanic stood still the entire time. There was an option to state when you wanted to exit. There was a paper introduction describing the situation in more detail. This simulation asked for the phone number.

Violations Found

1. H1: Visibility of System Status / Severity 4 / Found by: C

When the VR mechanic speaks, the only audio occurs. When the mechanic stops speaking, everything else is visually the same. The only difference is that there is not more audio. Thus, when the mechanic inserts pauses into his speech, I've talked over him multiple times because I cannot tell when I'm supposed to start and stop speaking (aka., a system status). Fixing this could be done by animating the mechanic's face to signal when the mechanic is and isn't talking.

2. H1: Visibility of system status / Severity 3 / Found by: A

Initial responses to my conversations had a rather long delay that I didn't know whether it was a technical error or I didn't say the proper words or otherwise. Perhaps acknowledging my words with an audio cue would help this disconnect.

3. H2: Match between system and the real world / Severity 2 / Found by: B, C

Although there are cars in the VR environment, since you can see the entire complex, it would be useful to actually be able to see your car which is confirmed to have already been brought in and is a Toyota. This would make the scene more realistic especially since the mechanic says the user's car is already there.

Fix: Add a graphic to the VR environment of the car the user owns that needs to be fixed.

4. H2: Match between system and the real world / Severity 1 / Found by: B

There is no microphone and for those who know how the VR headset works, will be confused how the user's voice is getting interpreted.

Fix: Add a mic component to the headset or on the desktop for the experiments.

5. H2: Match between system and the real world / Severity 2 / Found by: B, D

Both options of the mechanic's avatar have a monotone voice.

Fix: Choose more realistic voice with fluctuations to more mimic the actual experience someone would have.

6. H2: Match between system and the real world / Severity 2 / Seen by: A, B, C, D

The mechanic does not hand over the keys when he says "here are the keys."

Fix: Add motion of retrieving and handing over keys during the acquisition of the rental car portion of the experience.

7. H2: Match between system and the real world / Severity 1 / Seen by: A, B

The mechanic claims the rental car which is a Honda CRV is the front, but the car is not represented in what the user sees in VR.

Fix: Add graphic of CRV rental car.

8. H3: User control and freedom / Severity 4 / Found by: B

There is no way to exit the VR experience.

Fix: Add a voice command or button to exit the experience. Would be nice to be able to fully quit or save progress.

9. H4: Consistency and standards / Severity 3 / Found by: B

In the VR experience, at times the zombie-like mechanic's mouth moves although you do not hear him saying anything.

Fix: Only have avatar's mouth move when he is speaking to the user.

10. H4: Consistency & Standards / Severity 3 / Found by: A, B

The first option for the mechanic is an avatar with a very broad stance in a zombie like position. He constantly sways back and forth, has rips in his clothing, and no shoes.

This avatar is not particularly representative of a mechanic and resembles more of a zombie apocalypse-like character which might confuse users as to the purpose of the app. In the second version, the mechanic appears to constantly be dancing.

Fix: Change avatar to more typical mechanic persona.

11. H4: Consistency and Standards / Severity 2 / Found by: B

The mechanic knows details about the smaller types of rental cars, but doesn't know what a CRV is. To make the experience more realistic, the mechanic should be very knowledgeable about all the available rental cars.

Fix: Add data about rental cars to be used by the mechanic.

12. H4: Consistency and standards / Severity 1 / Found by: A, B

In the version with the dancing mechanic avatar, the goals box has inconsistent bullet point representation. Some items have a double dash while others have a single.

Fix: Make each goal displayed by the same representation of a bullet point.

13. H4: Consistency and Standards / Severity 2 / Found by: C

Considering that most people don't directly talk to a mechanic at a dealership, it is kind of weird that the simulation would start right at the discussion with the mechanic. This violates consistency between the system and the real world (also known as immersion in VR heuristics). I would include an introductory part where the person has to arrive at the dealership and introduce themselves before getting to the mechanic.

14. H4: Consistency and Standards / Severity 1 / Found by: C

Although the situation says that the user is at a dealership, the visuals look closer to a car garage because the only building in the near vicinity is the garage. This breaks consistency between the system and the real world because people expect to see cars being sold and other dealership personnel along with a car garage at a dealership. In fact, I was confused if I was at a dealership or not because of these visuals, which

caused me to wonder if my warranty would work, messing with my performance in the simulation. To fix this, I would include visuals associated with car dealerships, such as cars being sold along with explicit car (i.e., Toyota) branding.

15. H4: Consistency and Standards / Severity 1 / Found by: C

When the mechanic stated that my car's engine was overheated, there were no indications of where to look. Hence, I was confused, since the mechanic faces me, but my car is on my left side. Thus, this violates consistency between the system and real world because mechanics in the real world show where the problems are. I would add an animation of the mechanic pointing at the toast engine to fix this problem.

16. H5: Error prevention / Severity 2 / Found by: B, D

When the user says something incorrect, the mechanic just continues with the conversation.

Fix: Highlight the important piece of important from the notes or goals section.

17. H5: Error prevention / Severity 3 / Found by: A

After putting on the headset, I was immediately met with the mechanic figure and green floating screens; I could foresee a case where I'd feel pressured to read as quickly as possible and start speaking. Perhaps adding an area before being in the auto shop where the green screens follow us would help us feel like we purposefully entered this conversation.

18. H5: Error prevention / Severity 2 / Found by: A

Asking whether I needed a sedan or SUV was a weird question in that it came out of the blue and felt disjointed from the rest of the questions. Perhaps show both types of cars while out on the lot?

19. H6: Recognition rather than Recall / Severity 2 / Found by: A, C

Since I was given a prompt outside of the VR system that contained details not included in the system itself, there were some small details that didn't get included that I forgot about. Thus, there was a little bit of recall going on, violating the heuristic. It would have been nice to have just the prompter within the system only.

20. H6: Recognition rather than recall / Severity 4 / Found by: A

I found myself referring back to the list of things I needed to say often, but it was hard keeping track of what I'd already said and what I didn't because the information stayed constant. Animate this by crossing out tasks as they are completed in the system.

21. H7: Visibility of system status / Severity 4 / Seen by: A, B, C, D

Unsure what the actual end goal of the experience is and what a successful completion actually means. I was told that my goal is to get the car fixed, but I wasn't sure what truly meant I had completed the task.

Fix: Tell user when they have accomplished goal with either text or a badge system.

22. H7: Flexibility and efficiency of use / Severity 1 / Found by: B, D

The two boxes, goals and notes, both have a green background which can be hard to see for those who are color blind.

Fix: Change to color easily viewed or have option to choose color.

23. H7: Flexibility and efficiency of use / Severity 2 / Found by: B, D

The entirety of the notes and goals section remain for the whole experience which can be overwhelming to the user and not necessary if no longer needed.

Fix: Gray out pieces of information from notes and goals section as no longer needed which would also help the user know what is left for them to accomplish.

24. H7: Flexibility and efficiency of use / Severity 3 / Found by: A, C

When the mechanic asked about what happened to my car, I stated everything included on the information sheet. However, the mechanic started asking for information I had already provided in broken down form, although I had already provided this information. It violates flexibility and efficiency of use because I had to repeat what I had already provided again. A way to mitigate this issue is by providing additional branching features such that some questions can be skipped if the information has been provided (to prevent repetitive questions.)

25. H8: Aesthetic and minimalist design / Severity 2 / Found by: B

In the version with the notes section, the user's phone number is displayed. This section is purposefully visible in the native language, but since numbers are consistent, it takes up unnecessary room and was never required in the simulation.

Fix: Ensure that in all simulations, the phone number is asked for (inconsistency in the versions we saw).

26. H8: Aesthetic and minimalist design / Severity 1 / Found by: B

The trash can is knocked over and next to the car "getting fixed" in the garage during the experience. This may take away from the validity of the experience by making the garage look unsafe and untrustworthy.

Fix: Remove the trashcan and add tools someone would typically find in a garage to make the experience more realistic.

27. H8: Aesthetic and minimalist design / Severity 1 / Found by: B

The car in the garage behind the mechanic appears to be floating.

Fix: Lower the car in the scene or place it on a propped up surface.

28. H8: Aesthetic and minimalist design / Severity 3 / Found by: A, B

In the version with the mechanic wearing shoes, when the user looks down through the VR glasses, you appear to be standing partially inside the car.

Fix: Move car farther away in scene.

29. H8: Aesthetic and minimalist design / Severity 3 / Found by: A

The two green screens to the right appear to be flipped; I would expect my goals in interacting with the mechanic to be closer to eye level than background information, which I would likely refer to less than the goals. Switch the two content blocks.

30. H8: Aesthetic and minimalist design / Severity 1 / Found by: C

There were a lot of extra cars that were neither relevant to feel of the dealership nor the simulation. This violates (very minorly) the heuristic that one should include only necessary items. Thus, I would remove these cars since it doesn't add to the feeling of the dealership.

31. H8: Aesthetic and minimalist design / Severity 2 / Found by: A

Where are we? What is the location of this auto shop? I understand that the interaction is principally the most important part of the prototype, but the fact the shop exists in an oddly clean plane detracts my attention from the conversation at hand.

32. H8: Aesthetic and minimalist design / Severity 3 / Found by: A, B

The current scene looks like a garage in the middle of a wasteland. This apocalyptic world does not seem realistic and makes the experience seem more like a video game than a simulation of a real life experience.

Fix: Add nearby buildings (perhaps in the distance to not distract) and a road to show how the user could have even gotten to the garage.

33. H9: Help User with Errors / Severity 1 / Found by: C

When the user says something completely incomprehensible, the system does a good job recovering by asking the person to respond again via something similar to "I'm sorry, can you repeat that again?" However, most people will have forgotten the question in the first place, so it would be help users recover from errors by quickly restating the question. This way, the user doesn't have to ask what the question was again.

34. H9: Help User with Errors / Severity 3 / Found by: C, D

When the user says something completely contrary to their goal, the system effectively stops responding. (For more context, the goal of this simulation was to get my car engine fixed, but I tried a response that I didn't want my engine fixed. The simulation did not respond to this response.) This violates the heuristic that a system should help users recover from errors. Thus, a way to do this is by gently re-prompting the user of his/her goal (e.g., "I'm sorry, but don't you need to get your engine fixed?" or "Are you sure that you don't want to fix your engine? What are you going to do with your car?").

35. H10: Help and documentation / Severity 4 / Found by: A, B

There is no onboarding experience for users to learn what they are supposed to do or how they can interact with the environment.

Fix: Have an onboarding flow for new users to understand the goals and how to navigate through the experience as well as what success looks like.

36. H10: Help and documentation / Severity 4 / Found by: B, D

Unsure what to do as a user when I am absolutely lost in the experience due to language skills.

Fix: Incorporate a system of hints to help the user when struggling or a way to get guided through the English portion of the conversation.

37. H10: Help and documentation / Severity 3 / Seen by: B, C

It can be easy to get into a loop of confusion. For instance, when I ask what a coolant leak is, the mechanic responds with the engine is leaking coolant. I then asked what is coolant and it said it is coolant.

Fix: Add access to a dictionary or purposefully explain terms that a non-native speaker might not know.

38. H10: User control and freedom / Severity 2 / Found by: A

There lacked clear reference for if I was struggling to find a correct word. Add a toolbar or something of the sort I can hover with to get a translation dictionary.

39. S1: Give agent persona through language, sounds and other styles / Severity 1 / Found by: D

The robotic reading voice feels impersonal and unrealistic. Users may feel less immersed in the language-learning experience because the voice is distracting. Having actual voiced lines from a voice actor may help the user better process pronunciation of words and feel more attuned to the conversation.

40. S4: Start and stop conversations / Severity 3 / Found by D

There are no “wake words” for silences when a user has provided an incorrect response. Users may feel confused about the long silence and not be sure how to resume the conversation or whether to provide a different response. A solution would be to add voice lines for prompts (e.g. “Sorry, I didn’t catch that,” “I can’t find that model, did you mean something else?”)

41. S5: Pay attention to what the user said and respect their context / Severity 2 / Found by: D

The user doesn’t receive enough verbal feedback from the bot. This problem is twofold, since the user may not be sure whether their response was interpreted correctly, but also, as a language-learning application, the user is missing out on an opportunity to

practice reinforcement of language by hearing their response repeated. This could be resolved by repeating back what the user said in the speaker's voice.

42. S6: Use spoken language characteristics / Severity 1 / Found by: D

Similar to #8, the robotic voice doesn't feel too natural. The user may feel less connected to the conversation. The bot should use more language like "Got it" or "Hmm" or "Ahh" or other filler language tokens.

43. S14: Use multimodal feedback when available / Severity 3 / Found by: D

There were no visual cues when a user provide correct or incorrect feedback. The user may get lost in the situation without visual markers of what they have accomplished or still need to do. A solution to this would be to include flashing overlays when a user has provided appropriate responses as well as updating a status/progress bar that correlates with their progress in the scene.

Summary of Violations

Category	# Viol. (sev 0)	# Viol. (sev 1)	# Viol. (sev 2)	# Viol. (sev 3)	# Viol. (sev 4)	# Viol. (total)
H1: Visibility of Status	0	0	0	1	1	2
H2: Match Sys & World	0	2	3	0	0	5
H3: User Control	0	0	0	0	1	1
H4: Consistency	0	3	2	2	0	7
H5: Error Prevention	0	0	2	1	0	3
H6: Recognition not Recall	0	0	1	0	1	2
H7: Efficiency of Use	0	1	1	1	1	4
H8: Minimalist Design	0	3	2	3	0	8
H9: Help Users with Errors	0	1	0	1	0	2
H10: Documentation	0	0	1	1	2	4
Speech Heuristics	0	2	1	2	0	5
Total Violations by Severity	0	12	13	12	6	43

Evaluation Statistics

Severity/Evaluator	Evaluator A	Evaluator B	Evaluator C	Evaluator D	All evaluators
level 0	0%	0%	0%	0%	0% (0)
level 1	8% (1)	50% (6)	33% (4)	25% (3)	28% (12)
level 2	31% (4)	54% (7)	31% (4)	38% (5)	30% (13)
level 3	58% (7)	42% (5)	25% (3)	25% (3)	28% (12)
level 4	33% (2)	50% (3)	33% (2)	33% (2)	14% (6)
total (levels 3 & 4)	50% (9)	44% (8)	28% (5)	28% (5)	42% (18)
total (all levels)	33% (14)	49% (21)	30% (13)	30% (13)	100% (43)

Summary of Recommendations

Great job on this medium-fi prototype! As far as the user interface is, I think it's very intuitive to me as someone who's used VR; however, it'd be helpful if there were even an example interaction set up in user's native languages, since the tech is still relatively new and artificial and having that first interaction can help them acclimate to the tech.

Overall, you guys have a lot of functionality and a super interesting research question, but I think changing some of the details, especially visually, will take the prototype to the next level by making the experience more realistic. Specifically, adding a more realistic town around the auto shop and changing the mechanic's avatar would help showcase this experience as real vs. a video game.

It would also be helpful to have more support for confusion. When I first started the application, I was unsure of what to do and didn't realize I had to speak first. It would be great to have extra resources so users don't get in a sort of sinkhole that emphasizes the the language barrier further. I found this to be case when asking about what coolant is even though I was speaking in English. I could see a real use case for a dictionary as well as being able to pause the experience. In addition, it would be helpful to add more features that help the user know what to do next and when the task is complete. I wasn't sure of my status throughout the experience and I think the text boxes could be utilized really well to help the user know what to ask about next and when they have completed everything.

Also useful would be a thoughtful consideration of the presence of the green screens next to the participant when they're speaking to the AI. I think there's so much potential for the screens to supplement information like a translation dictionary, as well as a way to cross off objectives as they're completed; understandably, this all would be a lot to implement in a short time span, so take what I say here with a grain of salt.

I remember we also talked about making the AI nature of the conversational agent more salient and present; this was a great insight, as it helps participants be more forgiving of lag, delays, or minor faux pas in the conversation.