**Task: Telephonic Interactive Voice Response System with Voice Driven Graphical User Interface**

We are conducting a human study for the effectiveness of graphics an interactive voice response systems. We intend for voice driven graphics, even the text boxes and radio checkboxes have to be drivable by voice.

The main thing is:

There is a collections of *screens[..]* and a collection of *dialogue[..]* that flow together intertwined, i.e., if we interact with the screen and progress flow (ie, scene to scene), the dialogue interrupts and reverts to the next spoken line. If I barge-in and interact with the dialogue and progress flow, the screen is updated to the next screen. There must be spoken numbers and a haptic-enabled Dialpad with DTMF coded sounds.   
  
If I don't progress the flow, the system says "sorry" and repeats the same dialogue.  
  
Optionally, we can display the ASR output on screen and even highlight the text on the display.

It is not advised to flow in the text as the sound comes in, because the idea is that people will read faster than they listen or vice versa. So, the human interaction aspect must be well understood for the success of the system.  
  
*Expectations of the system:*

* Is Multiplatform, i.e., can work on phone, tablet and desktop natively and robustly.
* My thoughts on Implementations: Phone App, Web Dashboard, Desktop Client
* Due to time constraints, the telephonic line can be simulated rather than having to use Twilio.

Provided Context:

*IVR systems are phone based and allow customers and clients to call a company and automate various actions such as booking and appointment or asking for details of a delivery. Many companies are keen to move over to app or web-based systems to carry out the interaction as a visual element can often make the process faster and easier. However, a phone-based system is often required for customers and clients who are not comfortable with web based or app-based systems and also allow the option of speaking to a customer service representative if the problem is difficult to resolve.*

*In this project we explore the use of a multi-modal IVR system and a visual GUI style interaction which allows users to freely switch between both modalities.*

Provide Image:

A phone with a message on it

AI-generated content may be incorrect.

Attempting to put it on desktop, phone and tablet in 3 modes:

* GUI only [Transitions, Event handling, Screen Routing, Visual components, maybe even Virtual Avatar … but Mainly DTMF Dialpad, Chatbox, Haptic-involved Buttons]
* Voice Only [maybe ASR + LLM + Dialog Manager + TTS]
* **and GUI+Voice** (together with all the mechanisms and bells and whistles) [! Important]

This is so we can conduct a study to basically tell us that combining modalities makes it more effective if implemented the right way.

---

A developer’s sample attempt at prompting for the entire project as output on Claude Pro:  
  
A screenshot of a computer screen

AI-generated content may be incorrect.

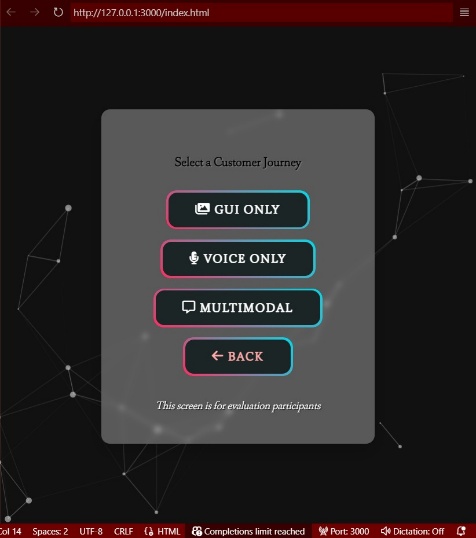
This method does not yielded working results, although there is a barrage of code files because of it.

**We strive for an industry-compliant attempt and not a quick throw-together.**

The following is the theme for the participant screens:

Please follow the theme guideline for the GUI and comply the flows to fit on a card with buttons:

A screenshot of a phone

AI-generated content may be incorrect. 

Please use a neural TTS capable of SSML.  
  
Attempt for Auto-generation of the GUI, Flow (in YML/TOML/JSON), and SSML  
  
Note:

For dialogue management, RASA is recommended.

For LLM, TinyLLaMa seemed of interest. Perhaps, we can distil an IVRLlama using Unsloth.

For TTS, Sesame was desired but because the API is not available yet, Hume’s Octave is fine option.

An architecture diagram with icons of the components used is required.

---

URGENT. IMMEDIATE ACTION REQUIRED.   
  
---  
  
TIME OVER – PROJECT IS RETIRED (7/4/25)