**IVR Modernization Document**

**1. Introduction**

An Interactive Voice Response (IVR) system is a telephony technology that allows callers to interact with an automated system using keypad inputs (DTMF) or simple voice prompts.

Traditionally, these systems are menu-driven,example:

“Press 1 for balance inquiry”

“Press 2 for recharge”

“Press 3 to talk to a customer care agent”

One of the most widely used technologies to build such systems is VoiceXML (VXML).

**2. What is VXML?**

VoiceXML (Voice eXtensible Markup Language) is an XML-based markup language designed specifically for creating voice-driven phone applications.

Similar to how HTML creates webpages, VXML creates voice-based menus for telephony.

It defines:

**Prompts** (messages played to the caller)

**User inputs** (key presses or speech recognition)

**Logic** to navigate menus

**3. How a VXML IVR Works**

**Steps:**

Caller dials IVR number

IVR plays menu options (using VXML prompts)

Caller responds (via keypad or voice input)

VXML logic processes input and directs the flow

Response/action is executed (fetch balance, transfer call, etc.)

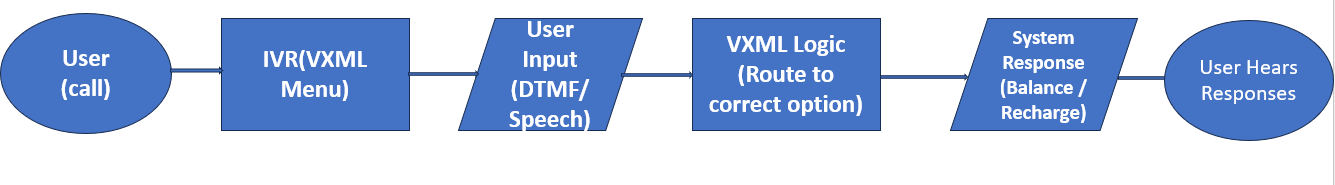
**4. Example Use Case**

A telecom operator’s VXML IVR might say:

“Welcome. Press 1 for Balance. Press 2 for Data Recharge. Press 3 to Speak with an Agent.”

If the user presses 1 → System fetches account balance → Plays back “Your balance is ₹450.”

1. **Traditional VXML IVR Flow (Diagram)**



1. **Example of Working with Middleware**

**6.1 Initial DTMF Interaction and Intent Mapping**

Middleware intercepts call and plays IVR prompt.

Example:

“For billing inquiries, press 1. For technical support, press 2. To speak with a representative, press 3.”

Customer presses ‘2’ → Middleware maps this DTMF tone to the “Technical Support” intent.

**6.2 Voice Stream Forwarding to NLP (ACS/BAP)**

Middleware activates voice input stage.

Customer says: “My internet connection is not working.”

Middleware forwards voice stream to NLP (e.g., ACS/BAP).

**6.3 NLP Processing (ACS/BAP)**

ASR: Converts speech to text.

NLU: Detects intent (“Troubleshooting Internet Connectivity”) and extracts entities (“internet connection,” “not working”).

**6.4 Response Generation**

NLP service determines next action:

Provide automated troubleshooting steps.

Create a support ticket.

Route to a technical support agent.

**Example:**

System responds: “Have you tried restarting your router?” (played back to user via middleware).

**7. Integration Needs (Old IVR ↔ ACS ↔ BAP)**

The old IVR (VXML-based) handles simple DTMF inputs and static prompts, while ACS + BAP enable natural language and AI-driven responses.

To integrate them, middleware is required.

**7.1 Key Integration Points**

**Input Handling**

**Old IVR:** DTMF / simple speech

**BAP:** Natural language (speech → intent)

**Need:** Middleware for DTMF → intent mapping + voice stream forwarding

**Output Handling**

**Old IVR**: VXML prompts

**BAP**: JSON responses

**Need**: Middleware for JSON ↔ VXML conversion

**Data Exchange Format**

**Old IVR**: XML (VXML)

**BAP/ACS**: JSON / APIs

**Need**: XML ↔ JSON transformation

**Session Management**

**Old IVR**: Menu position-based

**BAP:** Conversation state (intent + entities)

**Need:** Synchronize session IDs

**Backend Integration**

**Old IVR**: Direct backend calls

**BAP**: Middleware API calls

**Need:** Proxy backend access for both

**8. Example Modernized Flow**

User presses 1

IVR sends DTMF → Middleware converts into {“intent”: “CheckBalance”}

BAP fetches balance → returns {“message”: “Your balance is ₹450”}

Middleware converts into VXML → IVR plays back response

**9. Main Challenges in IVR Modernization**

**9.1 Real-time Response Limitations**

**Old IVR (VXML)**: Works with static menus and pre-recorded prompts.

**Conversational AI**: Generates dynamic responses instantly.

**Problem**: VXML may not handle real-time AI responses smoothly, causing delays or errors.

**9.2 Format Compatibility (VXML vs. JSON)**

**Old IVR**: Communicates using VXML (XML).

**BAP/ACS:** Communicate using APIs/JSON.

**Problem**: Continuous conversion XML ↔ JSON is required, introducing complexity and possible data loss.

**9.3 Session & Context Management**

**Old IVR:** Sessions are linear and menu-driven (tree-like: 1 → 2 → 3).

**BAP/ACS**: Sessions are contextual (remembers intent, entities, history).

**Problem**: Synchronizing old session IDs with AI context may be difficult. Conversations might feel broken if sync fails.

**9.4 Backend Integration Challenges**

**Old IVR**: Often directly calls backend systems (e.g., CRM for balance checks).

**BAP**: Expects API-based backend access.

**Problem**: Existing backends may not be easily compatible, requiring middleware bridging.