

Unit 8 · 4 Hours

VUI and NLP Based UI

Designing for voice assistants, conversational chatbots, game interfaces, and smartwatches.

BCA 2nd Semester · Tribhuvan University

Unit 8 Syllabus Roadmap

Four main topics covered in this unit:

Section	Topic	What You Will Learn
8.1	Command and Control vs Conversational UI	Differences between strict command menus and natural everyday conversation.
8.2	Personas, Avatars, Actors and Video Games	Character design, digital assistants, and the 4 types of game interfaces.
8.3	Speech Recognition & Dialog Management	How voice systems turn speech into actions, fill in details, and handle mistakes.
8.4	Designing for Wearable Devices	Designing for small watch screens, glanceability, and vibration feedback.

How User Interfaces Evolved

The way people interact with computers has gone through three big stages:

Stage 1

Command Line (CLI)

The user had to type exact code commands like `mkdir` or `rm`.

- Hard for beginners.
- You had to memorize exact words.
- Any typo produced an error.

Stage 2

Graphical UI (GUI)

The user clicks visual icons, buttons, and menus with a mouse or touch.

- Much easier to use.
- Options are visible on screen.
- You recognize choices instead of memorizing code.

Stage 3

Natural UI (VUI & NLP)

The user talks or types naturally, like talking to another person.

- No special syntax to learn.
- You state your goal in normal words.
- Works hands-free using voice.

Key Point: Early computers forced humans to speak like machines. Modern voice and chat systems allow computers to understand how humans naturally speak.

8.1 What is Command and Control UI?

Definition: A system where the user gives strict, exact commands that the computer executes step-by-step. The system only understands specific, predefined keywords or button presses.

Everyday Examples

- Terminal commands: Typing `git push` or `ping google.com`.
- Phone keypad menus: "Press 1 for English, Press 2 for Nepali, Press 3 for Support".
- Car dashboard buttons: Physical knobs to turn on the AC or switch radio channels.
- TV Remote: Separate buttons for volume, power, and channels.

Advantages & Disadvantages

- Fast for experts: A power user can type or press shortcuts very quickly.
- 100% predictable: The system does exactly what the button or command says.
- Hard to remember: Users must memorize exact words, flags, or button steps.
- Rigid: If you say or type the wrong word, the system fails with an error.

8.1 What is Conversational UI?

Definition: An interface where you interact using everyday human language, either through text chat or spoken voice.

1. Intent-Driven

The system focuses on what you want to do rather than how you phrase it.

Example: Saying "Send 500 rupees to Mom" or "Transfer 500 to my mother" achieves the same result.

2. Natural Dialogue

Interaction works like a normal two-way conversation.

If information is missing, the assistant asks a simple follow-up question instead of showing an error.

3. Context Memory

The system remembers what was said earlier in the conversation.

User: "Is it raining in Pokhara?"

Bot: "No, it's clear."

User: "How about tomorrow?" (Bot knows you mean Pokhara).

Popular Examples: ChatGPT, Google Assistant, Apple Siri, Amazon Alexa, and website support chatbots.

Comparing Command & Control vs. Conversational UI

A simple comparison table for exams and design decisions:

Feature	Command & Control	Conversational UI
How you interact	Type exact code words or press fixed buttons.	Speak or type in everyday human language.
What you must remember	High: you must memorize exact words and rules.	Low: you just describe what you want.
If you make a mistake	Fails immediately with an error message.	Asks a friendly question to clarify.
Flexibility	Rigid: you must follow strict steps in order.	Flexible: you can provide details in any order.
Best suited for	Expert tools, programming, car physical knobs.	Hands-free tasks, smart homes, customer help.

8.1 Multimodal UI: Combining Voice and Screen

Why Voice Alone Can Be Slow: Listening to a long list of 10 restaurant names over voice takes 40 seconds. On a screen, your eyes can scan all 10 options in 1 second.

Voice Input (Fast to Say)

- Speaking is 3 times faster than typing on a phone.
- Great when your hands are busy (cooking, driving, walking).
- Fast for simple requests: "Set timer for 10 minutes".

Visual Output (Fast to See)

- Screens show maps, photos, and comparisons at a glance.
- Information stays visible so you don't have to memorize it.
- Tap a button to confirm with one touch.

Everyday Examples: Google Maps in your car (you speak the destination, the screen shows the route) and smart displays like Amazon Echo Show or Google Nest Hub.

Interactive Demo: Command vs. Conversational UI

Try booking a flight to Pokhara using both styles to feel the difference:

Command & Control (Terminal)

Conversational UI (Chat)

Command & Control: Strict Syntax

Type 'help' for rules, or click one of the quick test buttons below:

Example: book -from KTM -to PKR -date 2026-10-15

Submit

Correct Syntax

Everyday English (Fails)

Missing Info (Fails)

Notice how the terminal requires exact flags like -from and -to, while chat understands your intent directly.

8.2 Personas in Voice and Chat

What is a Persona? The personality, tone of voice, and character of a voice assistant or chatbot. Even if you don't design one on purpose, users will naturally judge its personality.

Why Persona Matters

- People treat computers like people: Studies show users naturally expect politeness, empathy, and consistency from speaking agents.
- Matches the brand: A bank bot should sound trustworthy and clear. A game bot can be funny and playful.
- Sets expectations: If an assistant sounds like a genius human, people get very angry when it fails a simple question.

Design Guidelines

- Stay consistent: Don't switch between formal words and slang in the same conversation.
- Match the situation: Be cheerful when booking a vacation, but serious and polite when handling a cancelled flight.
- Avoid system codes: Say "I couldn't find that account" instead of "Error Code 404".

8.2 Avatars and the "Uncanny Valley"

What is an Avatar? A visual character or image that represents a user or a computer assistant (for example, the green owl in Duolingo, Apple Memoji, or a gaming character).

Why Use Avatars?

- Friendly face: Makes an app feel warm and welcoming.
- Visual cues: The character can smile, nod, or look thoughtful while processing.
- Habit building: Duolingo's mascot nudges you to practice daily.

The Uncanny Valley Rule

When a digital character looks almost human, but has slightly stiff movements or dead eyes, people feel creeped out.



Rule for Designers: Simple, cartoon-style avatars are almost always safer and friendlier than trying to make realistic human faces.

8.2 Actors in Interactive Systems

What is an Actor? Any person, bot, or program that has a role and can take actions in a system.

Type 1

Human User

The person giving goals, making decisions, and holding the final authority.

Type 2

Automated Agents

Software running in the background to help you: spam email filters, meeting schedulers, or price alert bots.

Type 3

Game Characters (NPCs)

Non-player characters in video games that talk, guide, or react to the player's choices.

Important Safety Rule: Whenever an automated bot is doing something important (like spending money or deleting files), it must always ask the human user for confirmation first.

8.2 The 4 Types of Game UI

How game interface elements are placed in the world and story:

1. Diegetic UI

In World + In Story

Part of the 3D game world and seen by the character.

Examples: A watch on the character's arm, car speedometer on the dashboard, or a map held in the character's hands.

2. Non-Diegetic UI

On Screen + Outside Story

A 2D overlay on the screen that only the player sees. The character cannot see it.

Examples: Health bar at the top of the screen, ammo count digits, or radar minimap.

3. Spatial UI

In World + Outside Story

Visual guides placed on the 3D floor or walls for the player. The character does not see them.

Examples: Glowing navigation arrows on a race track road, or an objective marker hovering over a door.

4. Meta UI

On Screen + In Story

Effects rendered on the screen representing the character's physical state.

Examples: Blood splatters on screen when shot, raindrops dripping down the camera lens, or blurred vision when dizzy.

8.2 Gamification: Using Game Ideas in Apps

Definition: Using game elements (points, streaks, progress bars) in regular apps like education, fitness, or banking to keep users motivated.

Common Game Elements

- Daily Streaks: Showing "Day 7 in a row" encourages users to return every day (e.g. Duolingo, Snapchat).
- Progress Bars: Showing "Profile 80% Complete" motivates people to finish the remaining steps (e.g. LinkedIn).
- Badges & Levels: Earning small rewards when you hit a goal (e.g. fitness apps).

Important Caution

- Points cannot fix bad design: If an app is confusing, adding badges will not save it.
- Must feel meaningful: Users care about real progress (learning a language, saving money), not meaningless numbers.
- Don't annoy users: Too many congratulation popups become irritating.

Interactive Demo: Game UI & Persona Tone

Game UI Inspector

Click each type to see a simple description and examples:

1. Diegetic

In World & Story

2. Non-Diegetic

Screen Overlay

3. Spatial

In 3D World Only

4. Meta

Camera Screen Effect

Diegetic UI

Part of the 3D game world and seen by the character.

Examples: A wristwatch on the character's arm, car speedometer on dashboard.

Assistant Tone Tuner

Adjust the sliders to see how tone changes the wording:

Formality: Casual ↔ Formal



Warmth: Direct ↔ Friendly



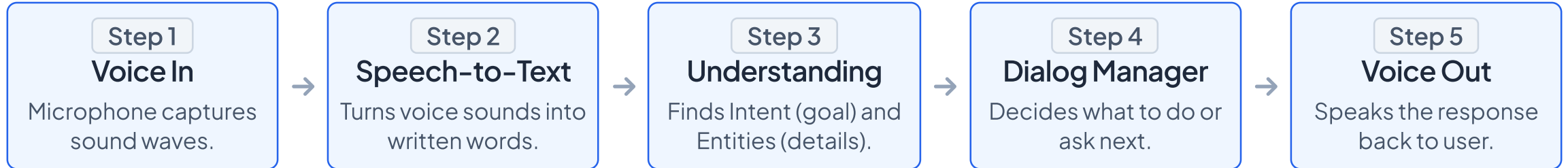
Tone Style:

Balanced

"You have NPR 42,500 available in your account. How can I help you today?"

8.3 Speech Recognition: How It Works

The 5 simple steps from user speaking to getting an answer:



Intent vs. Entity (In Plain English)

- Intent: What the user wants to do (the action).
Example: Book a ride, Set alarm, Check weather.
- Entity: The details needed to do it (the nouns).
Example: Destination: Thamel, Time: 7 AM.

The Dialog Manager

- Acts as the conversation brain.
- Checks which details have been provided and which are missing.
- Prompts for missing details before executing the task.

8.3 Why Speech Recognition is Challenging

Speech recognition works very well in quiet rooms, but real life is messy. Designers must plan for these four common problems:

1. Noise & Distance

- Background sounds: Traffic, kitchen fans, running water, TV playing in the room.
- Distance: Speaking to a smart speaker across the room makes the voice quiet and echoey.
- Multiple voices: Two people speaking at the same time confuses the microphone.

2. Language & Accents

- Accents: Non-native pronunciation (e.g. Nepali English) can be misunderstood by models trained on US voices.
- Words that sound identical: "two", "too", and "to" sound exactly the same.
- Mixing languages: Saying Nepali and English in one sentence ("Mero flight ko status check gara").

8.3 Dialog Management & "Slot Filling"

What is Slot Filling? Think of filling in the blanks of a form over voice. The system has empty slots for the details it needs. It only asks for the slots you haven't given yet.

Example: Booking a Flight

The system needs 3 slots: [From], [To], and [Date].

Slot 1: Destination
User said "Pokhara"

Filled

Slot 2: Travel Date
User said "Tomorrow"

Filled

Slot 3: Departure City
Not mentioned yet

Missing

System's Next Question: Instead of restarting from scratch, the bot asks only for the missing detail: "Where will you be flying from?"

8.3 4 Simple Rules for Voice Conversations

Paul Grice's conversational principles made simple for Voice UI:

Rule 1

Don't Say Too Much

Give just enough information, no more and no less.

Bad: Reading 4 paragraphs of flight history.

Good: "Your flight leaves at 8:00 AM."

Rule 2

Tell the Truth

Don't guess or make things up if you aren't sure.

If the microphone didn't hear clearly, admit it instead of doing the wrong task.

Rule 3

Stay on Topic

Directly answer what the user asked.

If the user asks "Do I need an umbrella?", say "Yes, rain is expected at 3 PM".

Rule 4

Be Clear and Brief

Put the most important answer first.

Say "Your balance is 5,000 rupees" rather than a long explanation first.

Main Takeaway: Write for the ear, not for the eye. Spoken messages must be short and direct because users cannot go back and re-read speech.

8.3 Handling Errors and Misunderstandings

Confirming Actions

- Ask Directly (for risky tasks): For sending money or buying tickets, ask explicitly:
"Send 500 rupees to Ram? Say yes to confirm."
- Mention Naturally (for safe tasks): For low-risk tasks, just repeat the detail while doing it:
"Playing Ed Sheeran on Spotify." (No need to stop and ask permission).

Progressive Help (3 Steps)

- 1st time it doesn't hear: Ask gently:
"Sorry, what city are you flying to?"
- 2nd time it fails: Give examples:
"You can say Pokhara, Biratnagar, or Chitwan."
- 3rd time it fails: Offer an alternative:
"I'm having trouble hearing. Let's use the screen instead."

Never repeat the same error message: Don't say "I didn't understand" 3 times in a row. Each time it fails, provide more helpful examples.

Interactive Demo: Speech & Slot Filling

Select a sample utterance to see how speech is converted to intent and slots:

- Sample 1: Pokhara Flight
- Sample 2: Doctor Reminder
- Sample 3: Weekend Weather

Recognized Speech

High (98%)

"Book a morning flight to Pokhara for tomorrow"

Goal (Intent): Book Flight

Details Extracted: Destination: Pokhara | Date: Tomorrow | Time: Morning

Slot-Filling Status

☐ Ask Explicit Confirmation

Destination: Pokhara

Date: Tomorrow

Departure City: Kathmandu (Default)

System Reply:

Implicit Confirmation: "Booking your morning flight to Pokhara for tomorrow. Which airline do you prefer?"

8.4 Designing for Wearable Devices

The 5–Second Rule: A smartwatch interaction should take no more than 2 to 5 seconds. If an action takes longer than 5 seconds, the user should pull out their smartphone instead.

Constraint 1

Tiny Screen

Smartwatch screens are around 1.5 inches. You cannot put complex tables, long paragraphs, or tiny buttons on them.

Constraint 2

Used While Moving

People check smartwatches while walking, jogging, or driving. Interactions must be readable with a quick glance.

Constraint 3

Small Battery

Batteries are small. Smartwatches use pure black backgrounds to save screen power and turn off pixels.

Core Rule: Smartwatches are for glancing, filtering, and quick taps; smartphones and laptops are for deep reading and typing.

8.4 How Users Interact with Smartwatches

The "Fat Finger" Problem

- Finger covers the screen: When your finger taps a 1.5-inch screen, it blocks almost half the display.
- Big buttons needed: Tap targets must be large (at least 7 to 9 mm) so users don't misclick while walking.
- No typing: Full QWERTY keyboards are frustrating on a watch. Use voice dictation or quick reply buttons instead.

Alternative Hardware Inputs

- Rotational Dial: Turning the Apple Watch crown or Samsung bezel scrolls without covering the screen with your finger.
- Wrist Gestures: Flicking your wrist or double-tapping fingers to answer calls or dismiss alerts.
- Voice input: Speaking a quick reply or asking Siri/Google Assistant bypasses the small screen completely.

8.4 Haptics: Feeling the Information

What is Haptic Feedback? Meaningful vibration patterns felt against the wrist. It allows the watch to communicate silently without demanding your eyes.

Why Haptics Are Useful

- Quiet and private: Alerts you in class or meetings without annoying ringtones.
- Eyes-free directions: When navigating on a bike or walking, 2 taps mean turn left; 3 taps mean turn right.
- Safety alarms: An abnormal heart rate spike triggers a strong, distinct vibration pattern.

Different Vibration Patterns

- Gentle single tap: Digital confirmation (e.g. payment completed at checkout).
- Double tap: Navigation turns or new incoming text message.
- Continuous pulse: Urgent phone call or health warning.

8.4 Organizing Information on a Watch

The three main layers of a smartwatch interface:

Layer 1

Watch Face & Complications

Tiny widgets on the main clock face: current battery %, step count, weather, or next calendar event.

Layer 2

Quick Notifications

Short alert cards that appear when a message arrives. Should have at most 1 or 2 action buttons ("Reply", "Dismiss").

Layer 3

Mini Apps

Simple scrolling tools when you deliberately open an app: workout tracker, timer, or music controller.

The 2-Tap Rule: Any action on a watch (replying to a text, pausing music, stopping a timer) should take no more than two taps.

8.4 Sensors and Privacy on Smartwatches

Sensors: Automatic Context

- Heart rate sensor: Continuously measures pulse and exercise intensity.
- Step and motion sensor: Tracks walking, running, and detects if the user falls down.
- Light sensor: Brightens the screen in direct sunlight and dims in a dark movie theater.

Wrist Privacy Challenges

- People around you can see: A sensitive bank alert or message popping up on your wrist in a crowd is visible to others.
- Privacy solution: The watch shows who sent the message, but hides the private text until you lift your wrist directly toward your face.

Proactive Help: Smartwatches don't wait for you to click. If you start running for 2 minutes, it asks: "Record outdoor run?"

Interactive Demo: Smartwatch & Haptics

Smartwatch Screen



Health

Weather

Battery

The 5-Second Glance Test

Click start, read the watch alert, and tap 'Dismiss' to test your reading speed:

Trigger Alert

Time: 0.0s

Vibration (Haptics) Tester

Click to feel/see the watch shake with different vibration patterns:

Turn Left

High Heart Rate

Payment Done

Turn Left

Two gentle taps telling you to turn without looking at the watch.

Summary: Comparing Screen, Voice, and Watch

Master summary of all interface types studied in this unit:

Interface	Input (Giving Info)	Output (Receiving Info)	Best Everyday Use
Screen (Desktop / Phone)	Keyboard typing or touch tapping.	Rich visual display (text, photos, maps).	Writing papers, reading news, editing photos, shopping.
Voice (VUI / Smart Speaker)	Fast speech in normal human language.	Audio voice (must be kept short).	Cooking, driving, playing music, smart home lights.
Chatbot (Text CUI)	Typing questions naturally.	Text messages and quick buttons.	Customer support, booking tickets, troubleshooting.
Wearable (Smartwatch)	Single taps, turning dial, wrist gestures.	Glanceable cards and vibration taps.	Health tracking, running, checking notifications in 3 seconds.

5 Simple Do's & Don'ts

5 Do's

- 1. Put the main answer first: Say the time or balance right away; don't make users wait through long introductions.
- 2. Offer examples when voice fails: If the bot doesn't hear, say "You can say Pokhara or Biratnagar".
- 3. Design for 2–5 second glances: Make smartwatch screens readable with a single quick look.
- 4. Use vibration for eyes-free alerts: Let users feel turns and alarms on their wrist.
- 5. Keep speech prompts short: Never read long paragraphs aloud.

5 Don'ts

- 1. Don't read long lists over voice: Never make users listen to 10 options without a screen.
- 2. Don't repeat identical errors: Never say "I didn't understand" 3 times in a row.
- 3. Don't shrink phone screens onto watches: Never copy-paste dense mobile layouts onto a 1.5-inch watch screen.
- 4. Don't force users to memorize code words: Let people speak in everyday language.
- 5. Don't make creepy realistic avatars: Simple stylized characters are friendlier and safer.

Tribhuvan University Exam Question Bank

Common questions asked from Unit 8 in BCA board examinations:

- Q1: Differentiate between Command & Control UI and Conversational UI with examples. [5 Marks]
- Q2: Explain the Speech Recognition pipeline (from voice input to voice output) and describe the role of the Dialog Manager. [10 Marks]
- Q3: Describe the 4 types of Video Game UI (Diegetic, Non-diegetic, Spatial, Meta) with gaming examples. [5 Marks]
- Q4: What are the key design constraints for wearable devices? Explain the 5-Second Rule and Haptic feedback. [5 Marks]
- Q5: What are the basic rules for designing voice user interfaces? Why is brevity important? [5 Marks]

Exam Tip: For Q2 (Speech Pipeline), always draw the 5-step block diagram (Voice In → Speech-to-Text → Understanding → Dialog Manager → Voice Out) and explain Intent vs Entity with a flight booking example!

Practice Exam Quiz

Test your understanding with 6 model questions:

Question 1 of 6

Score: 0 / 6

1. What is the main difference between Command & Control UI and Conversational UI?

A. Command & Control requires memorizing exact syntax; Conversational UI uses everyday language.

B. Command & Control uses voice only; Conversational UI uses typing only.

C. Conversational UI requires typing exact terminal codes.

D. There is no difference between them.

Practical Class Exercises

Three simple exercises to practice what you learned:

Exercise 1

Write a Voice Flow

Write a simple script for ordering a coffee by voice.

- Identify the Intent and Slots (Size, Sugar, Milk).
- Write the confirmation question.
- Write a polite message if it doesn't hear.

Exercise 2

Sketch a Watch Alert

Draw a smartwatch notification card on paper.

- Must be readable in 3 seconds.
- Include at most 2 big tap buttons.
- Use a dark background.

Exercise 3

Analyze a Game

Pick your favorite video game and find:

- One Diegetic element (in the story world).
- One Non-diegetic element (screen overlay).
- One Spatial or Meta element.

Unit 8 Completed

Great Job!

You now understand the essentials of Voice Interfaces, Conversational Systems, Game UI, and Wearable Design.

← Back to Unit 7

Restart Slides