Google AI Studio VOICE GENERATION

Complete A-to-Z Master Course

Zero Knowledge → Independent Voice Creator

🎯 প্রথমেই Course-এর একটি "Learning Promise"

> এই **course** এমনভাবে তৈরি করা হয়েছে যাতে আপনি আগে কখনও **Google AI Studio** ব্যবহার না করলেও, একেবারে **zero knowledge** থেকে শুরু করে ধাপে ধাপে **Google AI Studio** এবং **Gemini TTS-**এর **workflow** বুঝতে পারেন এবং নিজের প্রয়োজন অনুযায়ী **voice generation** করতে পারেন।

কিন্তু এখানে একটা গুরুত্বপূর্ণ educational honesty থাকবে:

> **AI-generated voice** সবসময় প্রথমবারেই **perfect** হবে—এমন **guarantee** নেই।
> **Professional result** পাওয়ার জন্য **output** শুনে **evaluate** করা এবং **prompt/script refine** করা এই **workflow-**এর স্বাভাবিক অংশ।

এটা খুব গুরুত্বপূর্ণ। কারণ learner-এর মধ্যে "আমি prompt দিলাম, কিন্তু perfect হলো না—মানে আমি ভুল করছি" এমন misconception তৈরি করা যাবে না। Google-এর নিজের prompting guidance-ও iteration-এর কথা বলে।

🧭 COURSE DESIGN PRINCIPLE

পুরো PPT-তে আমরা এই formula follow করব:

WHAT - এটা কী?

↓

WHY - এটা কেন দরকার?

↓

HOW - কীভাবে ব্যবহার করব?

↓

EXAMPLE - বাস্তবে কেমন?

↓

PRACTICE - নিজে করে দেখুন

↓

COMMON MISTAKE - মানুষ সাধারণত কোথায় ভুল করে?

↓

TROUBLESHOOT - ভুল হলে কী করবেন?

↓

REAL-WORLD APPLICATION - বাস্তবে কোথায় ব্যবহার করবেন?

এটাই পুরো এই Course-এর **teaching engine**।

PART 1 — AI সম্পর্কে Zero Knowledge থেকে শুরু

Chapter 1 — Artificial Intelligence কী?

একেবারে beginner-friendly definition।

তারপর:

Real-life examples

- Google Maps route suggestion
- YouTube recommendation
- Voice assistant
- AI chatbot
- Image generation
- Voice generation

তারপর explain:

> AI মানে “একটা magic box” নয়। বিভিন্ন AI model বিভিন্ন ধরনের কাজ করতে পারে।

Chapter 2 — Generative AI কী?

এখানে AI এবং Generative AI-এর পার্থক্য।

Traditional AI example

> “এই ছবিতে বিড়াল আছে কি?”

AI → উত্তর দেয়।

Generative AI

> “একটা Bengali children's story লিখে দাও।”

AI → নতুন content তৈরি করে।

তারপর:

Text Generation

Image Generation

Video Generation

Audio/Voice Generation

এই চারটা concept establish করা হবে।

PART 2 — Gemini Ecosystem

Chapter 3 — Gemini কী?

সহজ ভাষায়:

> Gemini হলো Google-এর generative AI model family/ecosystem, যার বিভিন্ন model বিভিন্ন ধরনের কাজের জন্য ব্যবহার করা যায়।

তারপর diagram:

Google

↓

Gemini

↓

Different capabilities/models

↓

Text / Image / Audio / Multimodal / etc.

Chapter 4 — Google AI Studio কী?

এটাই হবে প্রথম **major chapterI**

Official definition

Google AI Studio হলো Google-এর এমন একটি environment যেখানে Gemini models-এর সঙ্গে experiment, prompt testing, API key management এবং prototype/build করা যায়।

Beginner analogy

ধরো:

> **Google AI Studio = AI Workshop**

যেখানে তুমি AI-এর বিভিন্ন capability নিয়ে কাজ করে দেখতে পারো।

কিন্তু PPT-তে clearly লিখব:

> এটি একটি **teaching analogy; Google AI Studio-**এর **official definition** নয়।

এই distinction misinformation আটকাবে।

Chapter 5 — Google AI Studio কী নয়?

এটা খুব useful।

Google AI Studio:

❌ শুধু voice generator নয়
❌ শুধু chatbot নয়
❌ শুধু API console নয়
❌ শুধু developer-এর জন্য নয়

বরং এটি Gemini-based experimentation/prototyping-এর broader environment।

Chapter 6 — Google AI Studio দিয়ে কী করা যায়?

এখানে broad overview।

- Prompt experimentation
- Gemini interaction
- Multimodal work
- Prototyping
- App building
- API key management
- AI capability testing

Google-এর current AI Studio Build mode-এ natural-language prompt ব্যবহার করে application তৈরি ও modify করার workflow-ও আছে।

Chapter 7 — Google AI Studio বনাম Gemini

একটা comparison table:

বিষয়	Gemini	Google AI Studio
কী?	AI model family/ecosystem	AI working environment
মূল কাজ	AI capability	AI capability নিয়ে কাজ/test/build
Voice	Gemini-এর TTS capability থাকতে পারে	সেই capability test করার environment হতে পারে
Developer workflow	Model/API	Experiment/prototype/API management

---

PART 3 — এখন Voice Generation-এর দিকে যাওয়া

Chapter 8 — Audio কী?

শুধু এতটুকু:

> Audio হলো sound-এর recorded/generated form।

তারপর:

- Music
- Sound effects
- Speech
- Voice

Chapter 9 — Voice কী?

Voice হলো spoken human-like speech output।

Example:

> “আজকে আমরা AI Voice Generation শিখব।”

Chapter 10 — Speech কী?

Voice বনাম speech-এর difference খুব সহজভাবে।

Chapter 11 — TTS কী?

Text-to-Speech

Text → Spoken Audio

Example:

Text:

আজকে আবহাওয়াটা বেশ সুন্দর।

↓

TTS

↓

Audio:

[মানুষের মতো spoken voice]

Chapter 12 — Gemini TTS কী?

এখানে official definition:

> Gemini TTS হলো Gemini API-এর text-to-speech capability, যা text input থেকে single-speaker অথবা multi-speaker audio তৈরি করতে পারে এবং natural-language instruction দিয়ে speech-এর style, accent, pace এবং tone guide করা যায়।

এখানে অবশ্যই লিখব:

Current Status

Gemini TTS — Preview

কারণ official documentation বর্তমানে এটিকে Preview হিসেবে চিহ্নিত করছে।

Chapter 13 — Gemini TTS বনাম Live API

এটা বাদ দেওয়া যাবে না।

Gemini TTS	Live API
Scripted speech	Real-time interaction
Exact text recitation	Dynamic conversation
Style control	Interactive audio
Narration	Voice agent

Podcast/audiobook-type workflow	Real-time conversational experience

এই distinction Google-এর current documentation-এর সঙ্গে aligned।

PART 4 — Voice Generation-এর Basic Vocabulary

এখন learner-এর dictionary তৈরি হবে।

প্রতিটি term-এর জন্য:

Definition + Example + Visual

Speaker

কে কথা বলছে?

Transcript

কি বলা হবে?

Voice

কোন voice output ব্যবহার হবে?

Tone

কী ধরনের manner-এ বলা হবে?

Emotion

কী অনুভূতি থাকবে?

Pace

কত দ্রুত/ধীরে বলা হবে?

Accent

কী ধরনের pronunciation character থাকবে?

Style

Overall speaking character।

Context

কোন situation-এ বলা হচ্ছে?

PART 5 — প্রথমবার AI Studio ব্যবহার

এখানে PPT হবে **screen-by-screen tutorial।**

Chapter 14 — Google AI Studio Open করা

Step 1
AI Studio-তে যান।

Step 2
Sign in করুন।

Step 3
Relevant workspace/feature খুঁজুন।

এখানে screenshot থাকবে।

⚠️ Version Note

> UI সময়ের সঙ্গে পরিবর্তিত হতে পারে। Screenshot current version-এর reference; concept এবং workflow বেশি গুরুত্বপূর্ণ।

এই note অত্যন্ত জরুরি।

Chapter 15 — Voice Generation Workspace চিনে নেওয়া

Screenshot-এর উপর numbered labels:

① **Model**

② **Prompt/Input**

③ **Voice**

④ **Output**

⑤ **Controls**

ইত্যাদি।

যে UI option বর্তমান official documentation-এ নিশ্চিত নয়, সেটাকে course-এর permanent feature হিসেবে লেখা হবে না।

PART 6 — First Voice

Chapter 16 — "আমার প্রথম AI Voice"

প্রথম project হবে intentionally simple:

> "হাই! আমি আজকে Google AI Studio-তে আমার প্রথম AI voice তৈরি করছি।"

Step-by-step

1. Text দিন
2. Voice নির্বাচন করুন
3. Generate করুন
4. শুনুন
5. Note করুন

তারপর প্রশ্ন:

> Voiceটা কি natural লাগছে?

এখান থেকেই **quality evaluation mindset** তৈরি হবে।

PART 7 — Prompt-এর আসল শিক্ষা

Chapter 17 — Prompt কী?

Definition:

> AI-কে দেওয়া instruction/input।

Real-life analogy

> “আমার জন্য একটা চা বানাও।”

vs

> “এক কাপ গরম দুধ-চা বানাও, sugar কম, একটু strong করে।”

দ্বিতীয় instruction বেশি specific।

Chapter 18 — Voice Prompt কী?

Text + performance direction।

Example:

> Say this naturally in a friendly conversational tone.

তারপর transcript।

Chapter 19 — Weak vs Better vs Advanced Prompt

Level 1 : Say this.

Level 2 : Say this in a friendly Bengali voice.

Level 3 : Speak like a young Bengali man talking casually to his best friend. Keep the delivery natural, relaxed and conversational.

Level 4 : Character + Scene + Director's Notes + Transcript

Google-এর TTS prompting guide এই "director" mindset এবং Audio Profile, Scene, Director's Notes-এর মতো components-এর উপর জোর দেয়।

PART 8 — Prompt Structure Masterclass

এখানে আমরা একটা **Master Prompt Architecture** শেখাব।

[VOICE / AUDIO PROFILE]

[SCENE]

[CHARACTER PERSONALITY]

[DIRECTOR'S NOTES]

[STYLE]

[TONE]

[EMOTION]

[PACE]

[ACCENT / PRONUNCIATION]

[TRANSCRIPT]

[AUDIO TAGS]

কিন্তু Course -তে clearly লেখা থাকবে:

> এটি **Google-**এর বাধ্যতামূলক **syntax** নয়। এটি এই **course-**এর একটি **reusable teaching framework**। এই distinction misinformation আটকাবে।

PART 9 — Audio Profile

Character-এর identity:

Young Bengali male
Mid-20s
Friendly
Slightly playful
Soft voice
Casual conversational delivery

PART 10 — Scene

Example:

Rahul is sitting with his best friend at a roadside tea stall in Kolkata in the evening.

তারপর explain:

Scene → mood → performance

PART 11 — Director's Notes

এখানে:

- Speak naturally
- Keep the pace moderate
- Use subtle emotion
- Avoid sounding like a news presenter
- Pause briefly before important words

Google-এর official guide-এ style, accent, pace এবং tone-এর মতো performance characteristics direct করার কথা বলা হয়েছে।

PART 12 — Transcript

এখানে learner শিখবে:

Transcript = যা সত্যিই বলা হবে।

এবং:

Instruction ≠ Transcript

এই distinction খুব পরিষ্কার করা হবে।

PART 13 — Bengali Voice Masterclass

এখানে huge practical section থাকবে।

Written Bengali

> “তুমি কোথায় যাচ্ছ?”

Conversational Bengali

> “তুই কোথায় যাচ্ছিস?”

Bengali + English

> “দেখ, actually ব্যাপারটা একটু different।”

> “Wait, তুই seriously এটা করেছিস?”

এখানে শেখানো হবে:

Natural speech ≠ বইয়ের ভাষা।

তবে এটাও বলা হবে:

> Natural Bengali মানে ইচ্ছেমতো spelling বদলে দেওয়া নয়। Script-এর readability এবং intended pronunciation দুটোই বিবেচনা করতে হবে।

PART 14 — Emotion

প্রতিটি emotion আলাদা lesson।

Happy

Sad

Angry

Excited

Nervous

Surprised

Serious

Calm

Playful

Sarcastic

প্রতিটিতে:

Meaning → Scene → Prompt → Transcript → Expected delivery → Common mistake

---

PART 15 — Audio Tags

Google-এর current TTS documentation `[whispers]`, `[laughs]`, `[sighs]`, `[gasp]`, `[cough]`-এর মতো inline audio tags-এর examples দেয়। তবে Google একই সঙ্গে বলে exhaustive list নেই এবং experimentation করতে।

তাই PPT-তে লিখব:

> এই **tags guaranteed command-**এর মতো **treat** করা উচিত নয়**; model/version/context** অনুযায়ী **output** পরিবর্তিত হতে পারে।

এটা খুব গুরুত্বপূর্ণ misinformation control।

PART 16 — Pace, Pause & Rhythm

Example:

Normal

> “আমি দরজাটা খুললাম আর ভিতরে ঢুকে গেলাম।”

Suspense

> “আমি দরজাটা খুললাম...”

pause

> “তারপর ভিতরে ঢুকলাম।”

এখানে learner বুঝবে punctuation এবং explicit direction দুটোই performance-কে influence করতে পারে, কিন্তু output deterministic নয়।

PART 17 — Voice Selection

এখানে current Voice Library workflow শেখানো হবে।

Google-এর documentation AI Studio-র Voice Library-কে Gemini TTS-এর বিভিন্ন voice এবং speaking style পরীক্ষা করার জায়গা হিসেবে উল্লেখ করে।

Selection framework

Voice শুনে প্রশ্ন করো:

- আমার character-এর সঙ্গে match করছে?
- Age impression ঠিক?
- Energy ঠিক?
- Narration-এর জন্য suitable?
- Dialogue-এর জন্য suitable?
- Emotion carry করছে?

“এই voice best” বলা হবে না।

PART 18 — Character Voice Design

একটা complete **Character Voice Card**:

Character Name:

Age:

Personality:

Voice Type:

Energy:

Speaking Style:

Accent:

Typical Emotion:

Pace:

Relationship:

PART 19 — Voice Consistency

এখানে একটি critical truth:

> একই **prompt** দিলেই প্রতিবার **identical performance guaranteed** নয়।

তাই consistency-এর জন্য:

- stable character profile
- stable voice choice
- stable style instructions
- controlled script
- iterative review

ব্যবহার করতে হবে।

PART 20 — Multi-Speaker

Google TTS single-speaker এবং multi-speaker generation support করে।

Example:

> Speaker 1: “তুই কোথায় যাচ্ছিস?”
>
> Speaker 2: “বাজারে যাচ্ছি।”

তারপর individual direction:

> Speaker 1 — curious
> Speaker 2 — casual

PART 21 — Dialogue Acting

এখানে শুধু voice নয়:

Acting

শেখানো হবে।

Example:

> “তুই সত্যিই এটা করেছিস?”

একই sentence:

- shocked
- angry
- joking
- confused

PART 22 — Storytelling

Full story workflow:

Narrator Profile

↓

Scene

↓

Emotion

↓

Pacing

↓

Transcript

↓

Tags

↓

Generate

↓

Review

PART 23 — Real-World Projects

এখানে 10টি guided project:

Project 01 - Simple narration

Project 02 - YouTube explainer

Project 03 - Facebook Reel

Project 04 - Bengali story

Project 05 - Horror story

Project 06 - Two-person conversation

Project 07 - Animation dialogue

Project 08 - Advertisement

Project 09 - Podcast intro

Project 10 - 3-minute complete production

PART 24 — Voice Quality Control

প্রতিটি output-এর জন্য learner এই checklist ব্যবহার করবে:

Naturalness - মানুষের মতো শোনাচ্ছে?

Clarity - শব্দ পরিষ্কার?

Emotion - Scene-এর সঙ্গে match করছে?

Pace - Too fast / too slow?

Pronunciation - কোনো problematic word আছে?

Character - Personality consistent?

Acting - শুধু text পড়ছে, নাকি performance করছে?

PART 25 — Troubleshooting Bible

এটা Course-এর সবচেয়ে practical অংশগুলোর একটি হবে।

সমস্যা	সম্ভাব্য কারণ	কী পরীক্ষা করবেন
Robotic	context/direction দুর্বল	conversational direction
Too dramatic	emotion বেশি	subtle/minimal direction
Too fast	pacing unsuitable	slower pacing
Pronunciation issue	script/context	wording/context
Character mismatch	profile inconsistent	character profile
Wrong emotional feel	scene unclear	scene + emotion
Dialogue unnatural	written-style script	conversational rewrite

এখানে অবশ্যই বলা হবে:

> এগুলো **deterministic rules** নয়; **troubleshooting hypotheses**।

কারণ model output probabilistic/variable হতে পারে।

PART 26 — Prompt Iteration

এখানে Google-এর official principle সরাসরি শেখানো হবে:

> **Prompting is iterative.**

Google-এর prompting guide-এ clear instructions, context, examples এবং prompt iteration-এর উপর জোর দেওয়া হয়েছে।

Workflow:

Prompt 1

↓

Output শুনুন

↓

Problem identify করুন

↓

একটি variable change করুন

↓

Generate

↓

Compare

↓

Improve

PART 27 — Few-Shot / Example-Based Prompting

Google-এর general prompting guidance examples ব্যবহার করে desired pattern দেখানোর কথা বলে।

তাই advanced section-এ:

Example → Pattern → New Output

এই concept শেখানো হবে।

PART 28 — Advanced Prompt Architecture

এখানে learner একটি reusable professional template পাবে।

ROLE / CHARACTER

SCENE

PERSONALITY

VOICE CHARACTERISTICS

EMOTIONAL STATE

DELIVERY STYLE

PACE

ACCENT / PRONUNCIATION

PERFORMANCE NOTES

TRANSCRIPT

INLINE AUDIO DIRECTIONS

আবারও:

> এটি **course-created framework**, Google-এর mandatory syntax নয়।

PART 29 — API Track

এটা মূল beginner course-এর শেষে optional advanced module হবে।

এখানে:

- API কী?
- API key কী?
- AI Studio বনাম API
- authentication
- current API architecture
- Python

- JavaScript
- audio output
- automation

Google-এর বর্তমান documentation অনুযায়ী নতুন development-এর জন্য Interactions API হলো default interface; তাই outdated `generateContent`-কেন্দ্রিক workflow-কে main modern workflow হিসেবে শেখানো হবে না।

API key section-ও current authentication policy অনুযায়ী লেখা হবে। Google-এর documentation অনুযায়ী 2026 সালের মে থেকে নতুন keys auth keys হিসেবে তৈরি হচ্ছে এবং September 2026 থেকে standard keys নিয়ে migration requirement এসেছে।

PART 30 — Final Master Project

শেষে learner-কে একটি complete assignment:

🎬 "Create a 3-Minute Bengali AI Voice Production"

Requirements:

Character - 2–3 জন

Language - Bengali + natural English mix

Scene - Defined environment

Emotion - At least 3 emotional changes

Dialogue - Multi-speaker

Voice Direction - Character-specific

Audio Tags - Where appropriate

Quality Control - Self-review

Final Output - Production-ready audio