

Voice Notepad

User Manual v2

Application Version 1.8.0

PLATFORMS

Windows

Linux

Debian

Wayland

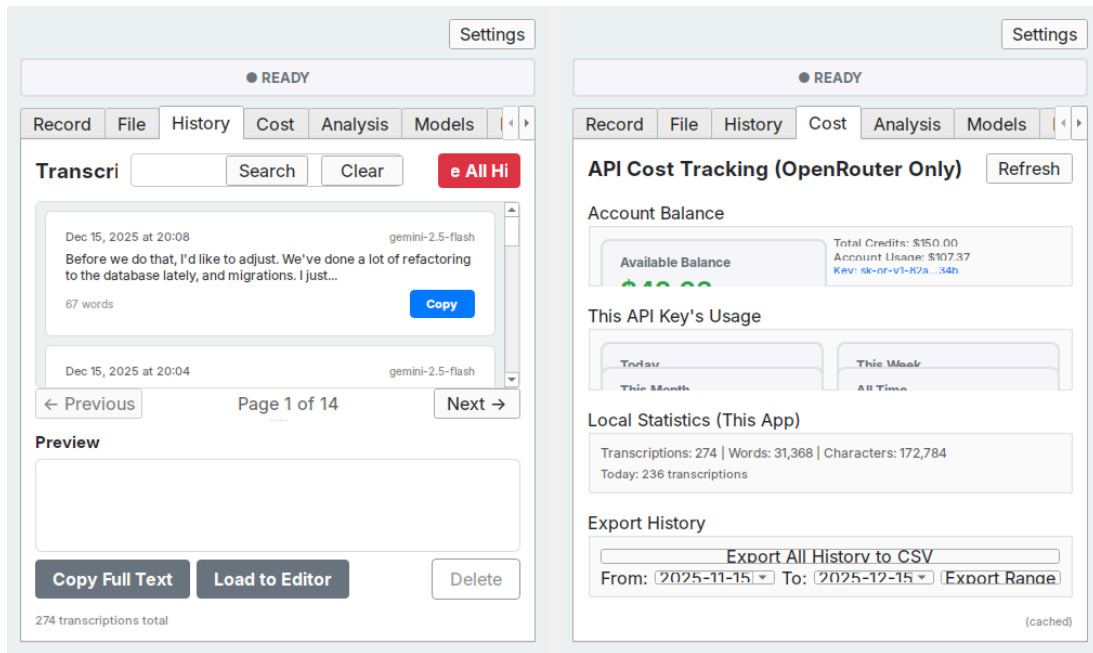
SUPPORTED AI PROVIDERS

OpenRouter

Google Gemini

OpenAI

Mistral AI



Author: Daniel Rosehill

Repository: github.com/danielrosehill/Voice-Notepad

License: MIT

Documentation created with Claude Code

Table of Contents

1	Introduction	5
1.1	Why Multimodal?	5
1.2	Key Features	5
2	Installation	6
2.1	Download Options	6
2.2	System Requirements	6
2.3	Linux Dependencies	6
2.4	Running from Source	6
3	Getting Started	8
3.1	Configure API Keys	8
3.2	Select Your Provider and Model	8
3.3	Start Recording	8
3.4	Copy or Save	8
4	Main Interface	9
4.1	Persistent Control Bar	9
4.2	Record Tab	10
4.3	File Transcription Tab	11
4.4	History Tab	11
4.5	Cost Tab	12
4.6	Prompt Stacks Tab	13
5	Settings	14
5.1	API Keys	14
5.2	Audio Settings	15
5.3	Behavior Settings	15
5.4	Personalization	16
5.5	Database	16
6	Format Presets	17
6.1	Foundational	17
6.2	Stylistic	17
6.3	Prompts	17
6.4	To-Do Lists	17
6.5	Documentation	17
6.6	Creative & Experimental	18
7	Global Hotkeys	19
7.1	Current Hotkey Mapping (F15-F19)	19
7.2	Workflows	19
7.3	Setting Up Hotkeys	19
7.4	Technical Notes	20
8	Keyboard Shortcuts	21
8.1	In-App Shortcuts	21

9	Audio Pipeline	22
9.1	Processing Stages	22
9.2	VAD Parameters	22
10	Cost Tracking	23
10.1	Cost Effectiveness	23
10.2	Status Bar	23
10.3	OpenRouter Tracking	23
11	Troubleshooting	24
11.1	No Audio Detected	24
11.2	Transcription Fails	24
11.3	High API Costs	24
11.4	Global Hotkeys Not Working	24
11.5	Text Injection Not Working	24
11.6	App Won't Start	25
12	Storage Locations	26
13	Support	27

Introduction

Voice Notepad is a desktop application for voice recording with AI-powered transcription and cleanup. Unlike traditional speech-to-text tools that require a separate text cleanup pass, Voice Notepad uses **multimodal AI models** that can process audio directly and perform both transcription and text cleanup in a single operation.

1.1 Why Multimodal?

Most voice-to-text apps use a two-step process: first transcribe with ASR (Automatic Speech Recognition), then clean up with an LLM (Large Language Model). Voice Notepad sends your audio directly to multimodal AI models that can hear and transcribe in a single pass.

This matters because:

- **The AI “hears” your tone, pauses, and emphasis** rather than just processing raw text
- **Verbal editing works naturally:** Say “scratch that” or “new paragraph” and the model understands
- **Faster turnaround:** One API call instead of two
- **Lower cost:** Single inference pass

1.2 Key Features

- **One-shot transcription + cleanup:** Audio is sent with a cleanup prompt to multimodal models
- **Multiple AI providers:** OpenRouter (recommended), Gemini, OpenAI, and Mistral
- **File transcription:** Upload audio files (MP3, WAV, OGG, M4A, FLAC)
- **Audio compression:** Automatic downsampling to reduce file size and API costs
- **Voice Activity Detection (VAD):** Optional silence removal before upload
- **Automatic Gain Control (AGC):** Normalizes audio levels for consistent results
- **Cost tracking:** Monitor your API spending in real-time
- **Transcript history:** All transcriptions are saved locally with full-text search
- **Global hotkeys:** System-wide shortcuts that work even when minimized
- **Prompt Stacks:** Layer multiple AI instructions for complex workflows
- **Email personalization:** Business and personal email signatures
- **Text injection:** Automatically paste transcriptions at cursor (Wayland)
- **Append mode:** Record multiple clips and combine before transcription

Installation

2.1 Download Options

Voice Notepad is available in multiple formats:

Format	Platform	Description
.exe	Windows	Windows installer
.zip	Windows	Portable Windows version
.AppImage	Linux	Universal Linux package (runs anywhere)
.deb	Debian/Ubuntu	Native Debian package
.tar.gz	Linux	Portable archive

Download the latest release from: <https://github.com/danielrosehill/Voice-Notepad/releases>

2.2 System Requirements

- **Operating System:** Windows 10+ or Linux
- **Python:** 3.10+ (for running from source)
- **Audio:** Working microphone
- **Internet:** Required for API calls

2.3 Linux Dependencies

```
sudo apt install python3 python3-venv ffmpeg portaudio19-dev
```

2.4 Running from Source

```
git clone https://github.com/danielrosehill/Voice-Notepad.git
cd Voice-Notepad
./run.sh
```

The script creates a virtual environment using `uv` and installs dependencies automatically.

Getting Started

3.1 Configure API Keys

Before using Voice Notepad, you need to set up at least one API key:

1. Click the **Settings** button in the top-right corner
2. Go to the **API Keys** tab
3. Enter your API key for your preferred provider

Recommended: Use **OpenRouter** for access to multiple models with a single API key and accurate cost tracking.

3.2 Select Your Provider and Model

On the Record tab:

1. Choose your **Provider** from the dropdown (OpenRouter, Gemini, OpenAI, Mistral)
2. Select your preferred **Model** from the model dropdown
3. Use **Standard** or **Budget** tier buttons for quick model switching

3.3 Start Recording

1. Click **Record** or press Ctrl+R
2. Speak into your microphone
3. Click **Transcribe** or press Ctrl+Return when finished
4. Your cleaned transcription will appear in the text area

3.4 Copy or Save

- Click **Copy** to copy to clipboard
- Click **Save** to save to a file
- Enable **auto-copy** in Settings to automatically copy after transcription

Main Interface

Voice Notepad uses a tabbed interface with several main sections. The recording controls are always visible at the top of the window across all tabs.

4.1 Persistent Control Bar

The control bar at the top of the window contains:

- **Status indicator:** Shows current state (Ready, Recording, Processing, etc.)
- **Duration timer:** Shows recording length
- **Segment count:** Shows number of cached segments in append mode
- **Record button:** Start a new recording
- **Pause button:** Pause/resume recording
- **Append button:** Record additional audio to combine with cached audio
- **Stop button:** Stop recording and cache audio
- **Transcribe button:** Send cached audio for transcription
- **Delete button:** Clear current recording and cached audio

4.2 Record Tab

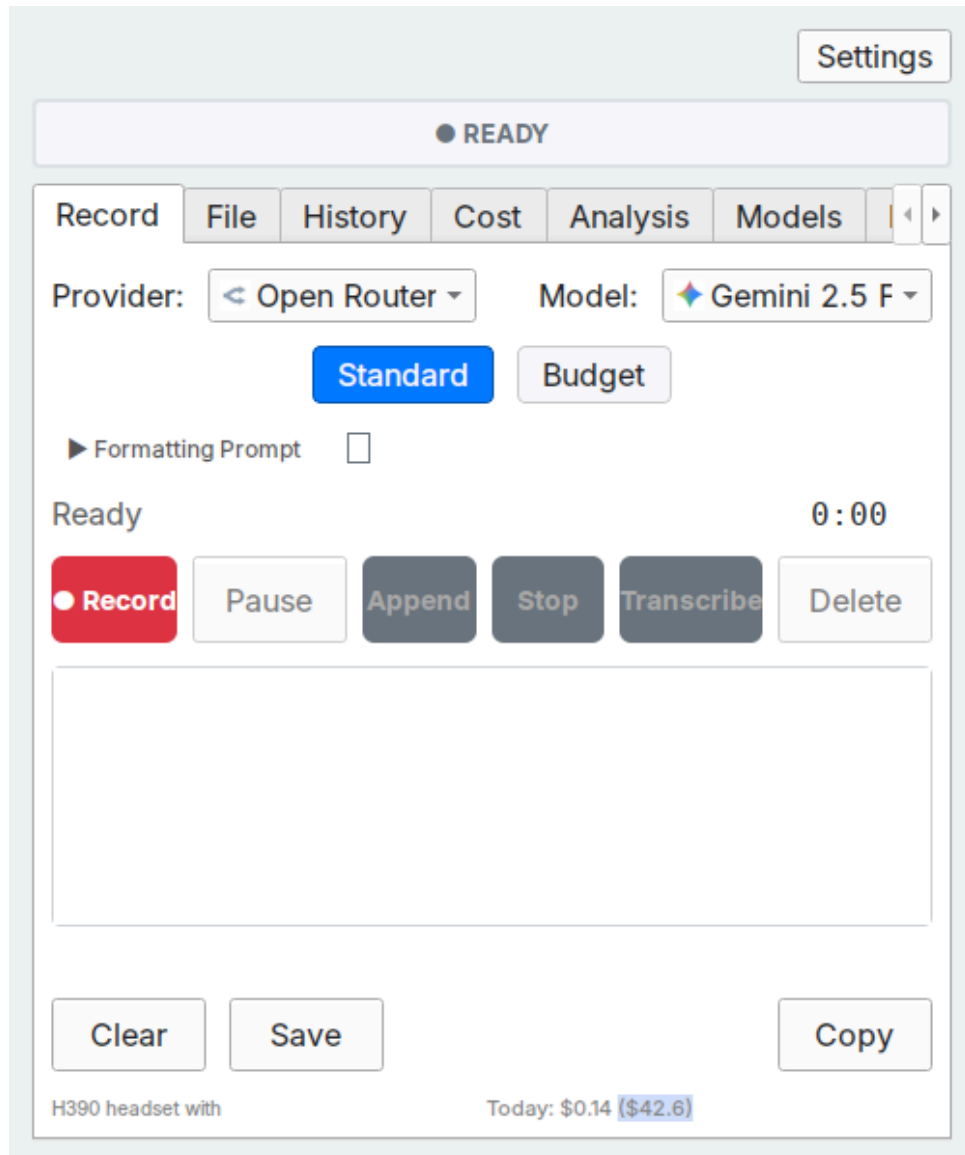


Figure 1: The Record tab - main transcription interface

The Record tab is where you perform transcriptions.

Model Selection:

- **Provider dropdown:** Select your AI provider
- **Model dropdown:** Choose the specific model
- **Standard/Budget buttons:** Quick tier switching

Favorites Bar:

- Quick access to your favorite format presets
- Configure favorites in Settings → Personalization

Prompt Controls:

- **Format preset:** Select output format (General, Email, Todo, etc.)
- **Formality:** Casual, Neutral, or Professional tone

- **Verbosity:** Control how concise the output should be

Output Area:

- View your transcribed text with markdown rendering
- Toggle between **Rendered** view and **Source** editing
- Word and character count displayed

Actions:

- **Clear:** Clear the text area
- **Save:** Save transcription to a file
- **Copy:** Copy to clipboard
- **Rewrite:** Send text back to AI for further editing

4.3 File Transcription Tab

Transcribe existing audio files without recording.

Supported Formats:

- MP3, WAV, OGG, M4A, FLAC

Features:

- Drag and drop files or click to browse
- Multiple file selection
- Same prompt options as recording
- All transcriptions saved to history

4.4 History Tab

The History tab provides access to all your past transcriptions.

Features:

- **Full-text search:** Search across all transcriptions
- **Click to preview:** Single-click shows preview
- **Double-click to load:** Load transcription into editor
- **Delete individual:** Remove specific transcriptions
- **Delete All History:** Clear entire history with confirmation

Statistics:

- Total transcription count
- Database size
- Audio archive size (if enabled)

4.5 Cost Tab

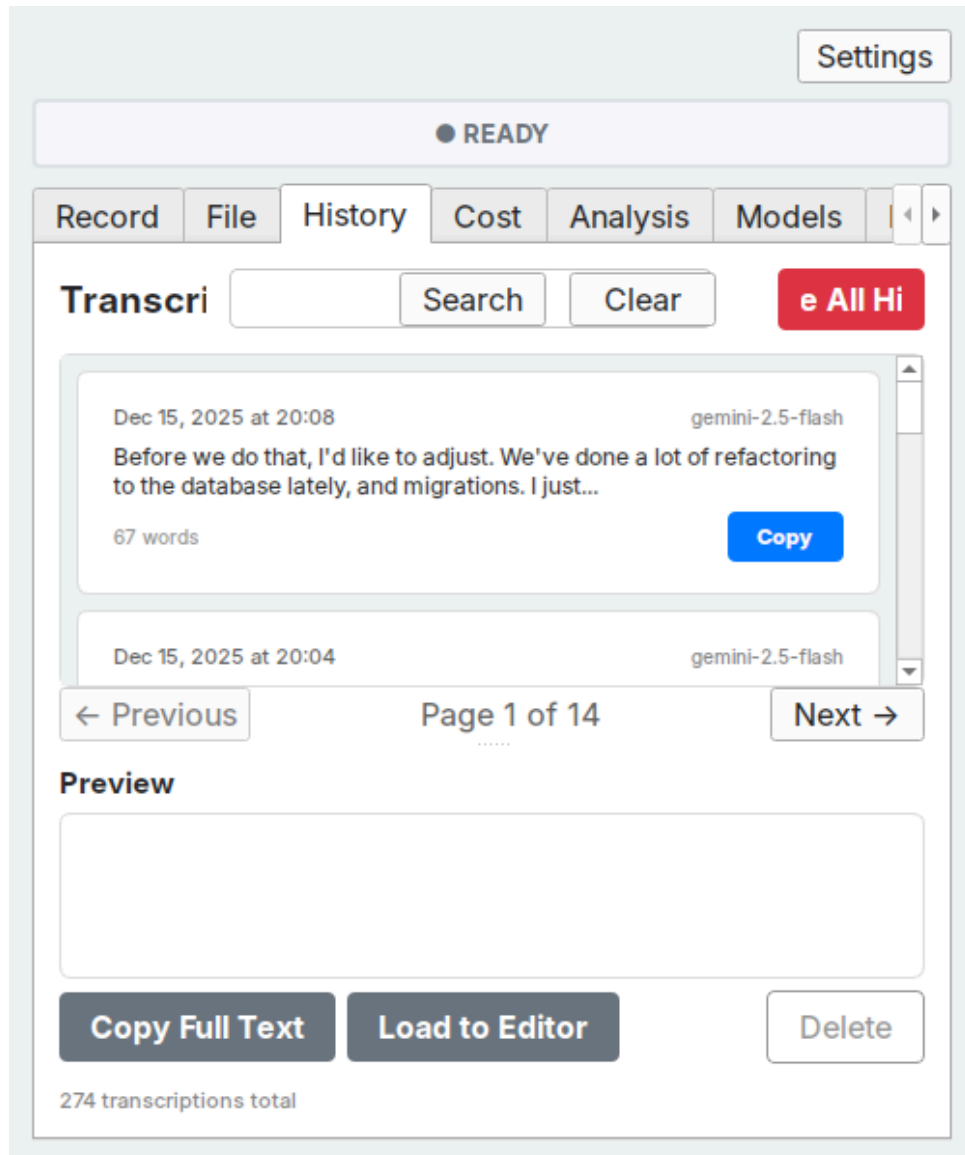


Figure 2: The Cost tab - API usage tracking

The Cost tab provides detailed API usage tracking.

Account Balance (OpenRouter only):

- Shows your available credit balance
- Displays total credits and usage

This API Key's Usage:

- **Today:** Current day spending
- **This Week:** Weekly spending total
- **This Month:** Monthly spending total
- **All Time:** Lifetime spending

Local Statistics:

- Transcription count

- Total words and characters processed

Export:

- Export history to CSV (all or date range)

4.6 Prompt Stacks Tab

Prompt Stacks allow you to layer multiple AI instructions for complex transcription scenarios.

Use Cases:

- Meeting notes with action items extraction
- Technical documentation with code formatting
- Multi-language transcription with translation
- Custom workflows for recurring tasks

Features:

- Create and save named prompt stacks
- Combine multiple prompts that stack together
- Import/export as JSON for sharing
- Toggle between legacy format system and prompt stacks

Settings

Access settings by clicking the **Settings** button in the top-right corner.

5.1 API Keys

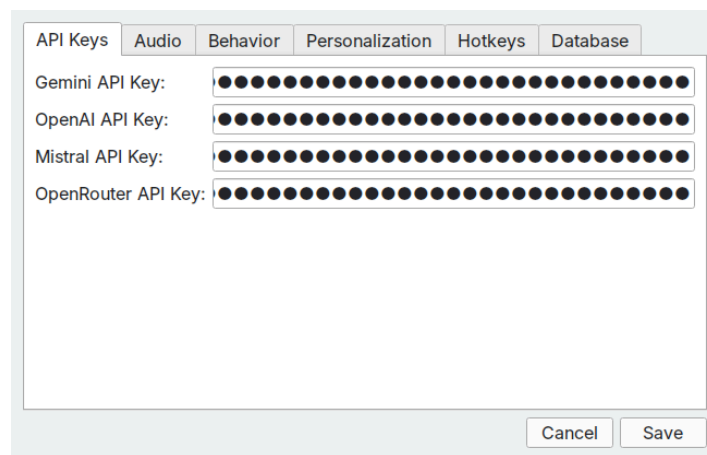


Figure 3: API Keys configuration

Configure your API keys for each provider:

- **OpenRouter** **API Key:** Unified multi-provider access (recommended)
- **Gemini** **API Key:** Direct Google AI access
- **OpenAI** **API Key:** Direct OpenAI access
- **Mistral** **API Key:** Direct Mistral access

API keys are stored locally in `~/.config/voice-notepad-v3/config.json`.

5.2 Audio Settings

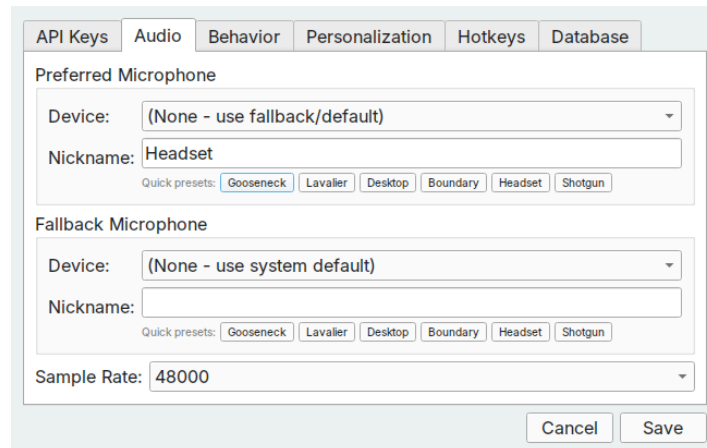


Figure 4: Audio settings configuration

Configure audio-related options:

- **Preferred microphone:** Select your primary input device with optional nickname
- **Fallback microphone:** Backup device if preferred is unavailable
- **Sample rate:** Audio quality settings (default: 48kHz)

5.3 Behavior Settings

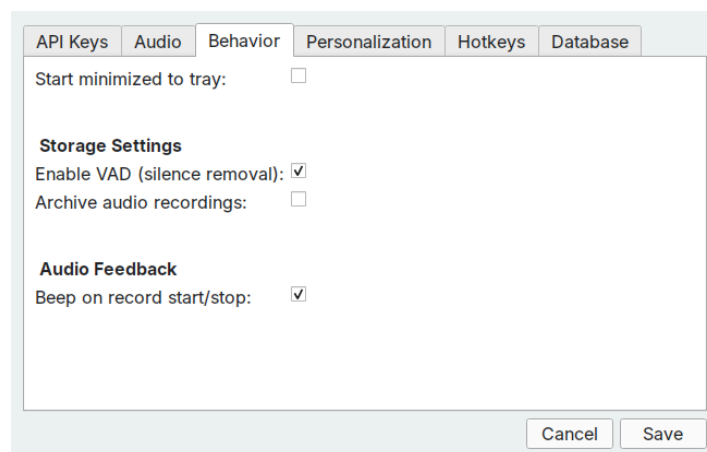


Figure 5: Behavior settings configuration

Customize app behavior:

- **Enable VAD:** Voice Activity Detection removes silence before upload
- **Enable AGC:** Automatic Gain Control normalizes audio levels
- **Auto-copy to clipboard:** Automatically copy transcriptions
- **Auto-paste (text injection):** Paste transcription at cursor after copy (requires ydotool on Wayland)
- **Archive audio:** Save Opus copies of recordings (24kbps)
- **Start minimized:** Start the app minimized to tray
- **Audio feedback:** Play beeps on recording start/stop

5.4 Personalization

Configure user-specific settings:

- **User name:** Your name for email signatures
- **Business email:** Email address for professional communications
- **Business signature:** Full signature block for business emails
- **Personal email:** Email address for personal communications
- **Personal signature:** Signature for personal emails
- **Favorite formats:** Quick-access format presets for the main UI

5.5 Database

Database management options:

- **Statistics:** Total transcriptions, database size, archive size
- **FTS Status:** Shows if Full-Text Search is enabled
- **Optimize Database:** Run VACUUM to reclaim disk space
- **Refresh Statistics:** Update the displayed numbers

Format Presets

Voice Notepad includes many format presets organized by category:

6.1 Foundational

- **General:** No specific formatting—general cleanup only
- **Verbatim:** Minimal transformation, closest to verbatim transcription

6.2 Stylistic

- **Email:** Professional email format with greeting and sign-off
- **Meeting Notes:** Structured notes with action items
- **Bullet Points:** Concise bullet point list
- **Internal Memo:** Company memo format
- **Press Release:** Corporate press release structure
- **Newsletter:** Email newsletter content

6.3 Prompts

- **AI Prompt:** General AI assistant instructions
- **Development Prompt:** Software development instructions
- **System Prompt:** AI system prompt (third-person)
- **Image Generation Prompt:** Prompts for AI image generators

6.4 To-Do Lists

- **To-Do:** Checkbox task list format
- **Shopping List:** Categorized shopping list
- **Grocery List:** Categorized grocery list

6.5 Documentation

- **Documentation:** Clear, structured documentation
- **Technical Docs:** Technical documentation and guides
- **README:** GitHub-style README

- **Reference Doc:** Quick-lookup reference material
- **API Documentation:** API endpoint documentation
- **SOP:** Standard Operating Procedure
- **Changelog:** Software release changelog

6.6 Creative & Experimental

- **Blog Post:** Full blog post with sections
- **Social Post:** Social media post
- **Story Notes:** Creative writing notes
- **Shakespearean Style:** Elizabethan English style
- **Pirate Speak:** Pirate vernacular

Global Hotkeys

Global hotkeys work system-wide, even when Voice Notepad is minimized or unfocused.

7.1 Current Hotkey Mapping (F15-F19)

Key	Action	Description
F15	Simple Toggle	Start recording, or stop and transcribe immediately
F16	Tap Toggle	Start recording, or stop and cache audio (for append mode)
F17	Transcribe Only	Transcribe cached audio without starting a new recording
F18	Clear/Delete	Delete current recording and clear all cached audio
F19	Append	Start a new recording that appends to cached audio

7.2 Workflows

Simple Workflow (F15 only):

1. Press **F15** to start recording
2. Press **F15** again to stop and transcribe

Append Workflow (F16/F17/F19):

1. Press **F16** to start recording
2. Press **F16** again to stop and cache (audio is held in memory)
3. Press **F19** to record another segment (appends to cache)
4. Press **F17** to transcribe all cached segments together
5. Press **F18** to clear cache and start over

7.3 Setting Up Hotkeys

Most keyboards don't have F15+ keys. Use **Input Remapper** to map other keys or buttons to these keycodes:

```
sudo apt install input-remapper
```

Common remapping options:

- **Pause/Break key** → F15 for toggle recording
- **USB foot pedal buttons** → F15/F17/F18 for hands-free operation
- **Extra mouse buttons** → F15 for quick dictation

7.4 Technical Notes

- On Wayland: Hotkeys work via evdev (reads directly from input devices)
- Requires user to be in the 'input' group for evdev access
- Falls back to pynput/X11 on non-Linux systems

Keyboard Shortcuts

8.1 In-App Shortcuts

These work when the Voice Notepad window is focused:

Shortcut	Action
Ctrl+R	Start recording
Ctrl+Space	Pause/Resume recording
Ctrl+Return	Stop and transcribe
Ctrl+S	Save to file
Ctrl+Shift+C	Copy to clipboard
Ctrl+N	Clear editor

Audio Pipeline

Voice Notepad processes audio through several stages before sending it to AI models.

9.1 Processing Stages

1. **Recording:** Captures at device's native sample rate (typically 48kHz)
2. **Automatic Gain Control (AGC):** Normalizes audio levels. Target peak: -3 dBFS, max gain: $+20$ dB
3. **Voice Activity Detection (VAD):** Silero VAD removes silence segments (typically 30-50% reduction)
4. **Compression:** Downsampled to 16kHz mono (matches Gemini's internal format)
5. **API Submission:** Audio is base64-encoded and sent with cleanup prompt
6. **Storage:** Transcripts saved to MongoDB-compatible database

9.2 VAD Parameters

Parameter	Value
Sample rate	16kHz
Window size	512 samples (32ms)
Threshold	0.5
Minimum speech	250ms
Minimum silence	100ms
Padding	30ms

Cost Tracking

Voice Notepad tracks API costs with **OpenRouter providing accurate key-specific costs** from the API.

10.1 Cost Effectiveness

Real usage data using Gemini 2.5 Flash:

- **848 transcriptions** for **\$1.17 total**
- **84,000 words** transcribed and cleaned
- About **\$0.014 per 1,000 words** (1.4 cents)

10.2 Status Bar

Shows “Today: \$X.XXXX (N) | Bal: \$X.XX” when using OpenRouter:

- N = number of transcriptions today
- Bal = remaining OpenRouter credit balance

10.3 OpenRouter Tracking

OpenRouter provides the most accurate cost tracking:

- **Key-specific usage:** Uses `/api/v1/key` endpoint
- **Account balance:** Via `/api/v1/credits`
- **Activity breakdown:** Model usage via `/api/v1/activity`

Troubleshooting

11.1 No Audio Detected

1. Check that your microphone is connected
2. Verify the correct microphone is selected in Settings → Audio
3. Check system audio settings (PipeWire/PulseAudio)
4. Try a different input device

11.2 Transcription Fails

1. Verify your API key is correct in Settings
2. Check your internet connection
3. Ensure you have credits/quota with your provider
4. Try a different provider or model

11.3 High API Costs

1. Enable VAD (Voice Activity Detection) to remove silence
2. Use Budget tier models when appropriate
3. Keep recordings concise
4. Monitor spending in the Cost tab

11.4 Global Hotkeys Not Working

On Linux/Wayland:

1. Ensure you're in the 'input' group: `sudo usermod -aG input $USER`
2. Log out and back in after adding to group
3. Check if input-remapper is properly configured

General:

1. Try different key combinations
2. Avoid keys that conflict with other applications

11.5 Text Injection Not Working

1. Ensure ydotool is installed: `sudo apt install ydotool`

2. Start the ydotool daemon: `sudo ydotoold &`
3. Check that auto-paste is enabled in Settings → Behavior

11.6 App Won't Start

1. Check system dependencies are installed
2. Verify Python version (3.10+)
3. Try running from terminal to see error messages
4. Reinstall the application

Storage Locations

All data is stored in `~/.config/voice-notepad-v3/`:

Path	Contents
<code>config.json</code>	API keys and preferences
<code>mongita/</code>	MongoDB-compatible transcript database
<code>usage/</code>	Daily cost tracking JSON files
<code>audio-archive/</code>	Opus audio recordings (if enabled)
<code>models/</code>	Downloaded VAD model (<code>silero_vad.onnx</code>)

Support

- **GitHub Issues:** <https://github.com/danielrosehill/Voice-Notepad/issues>
- **Documentation:** <https://github.com/danielrosehill/Voice-Notepad/tree/main/docs>

Voice Notepad User Manual v2 (Application v1.8.0)
Created by Daniel Rosehill & Claude Code
MIT License