

En Clair: A Portable Device for Making Perceptual Filtering Audible

Josh Camacho
Massachusetts Institute of Technology
Cambridge, MA, USA
joshua47@mit.edu

Elaine Liu
Massachusetts Institute of Technology
Cambridge, MA, USA
exliu@mit.edu

Krystal Montgomery
Massachusetts Institute of Technology
Cambridge, MA, USA
mkrystal@mit.edu

Abstract

Human communication is never lossless. What one person says and what another perceives are shaped by invisible filters — assumptions, emotional states, prior history, and cognitive biases — that neither party can directly observe. *En Clair* is a portable device that makes this filtering process audible and tangible. A listener carries the device during a conversation; speech directed at them is captured, transcribed, processed through an LLM conditioned by a user-defined perceptual filter, and played back through noise-cancelling headphones via text-to-speech. The listener hears not what was said, but what their filter makes of it. Filters are encoded on physical NFC cartridges programmed through natural language voice prompts; the same utterance sounds entirely different depending on which cartridge is loaded. Inspired by the Sony Walkman’s reframing of music as personal and portable, *En Clair* turns miscommunication into a designed object — one that surfaces the subjective, constructed nature of what we hear, and invites reflection on the filtering we do every day without noticing.

Keywords

perceptual filtering, large language models, tangible interaction, NFC, portable computing, miscommunication, human-computer interaction

ACM Reference Format:

Josh Camacho, Elaine Liu, and Krystal Montgomery. 2026. En Clair: A Portable Device for Making Perceptual Filtering Audible. In *Proceedings of 27th ACM International Conference on Multimodal Interaction (ICMI '25)*. ACM, New York, NY, USA, 2 pages.

1 Introduction

Every act of listening involves interpretation. Long before a word is consciously processed, the mind applies filters shaped by mood, relationship history, and cultural context that transform a speaker’s meaning into something that belongs, at least partly, to the listener. This filtering is normal, unavoidable, and almost entirely invisible.

En Clair — from the French *en clair*, “in plain text,” — names its aspiration ironically: the device’s output is anything but plain. It sits at the user’s waist during a conversation. The listener wears noise-cancelling headphones; the device captures the speaker’s voice,



Figure 1: En Clair open view

transcribes it via the OpenAI Whisper API, passes the text through Claude (Anthropic) with a cartridge-defined system prompt, and plays the transformed speech back through the headphones via ElevenLabs text-to-speech. The headphones do double duty: blocking the speaker’s actual voice while delivering the filtered version in its place.

Prior work has explored mediated communication as art and design material. Wright’s *Conversation Piece* [4] examined selective encoding in conversational memory; Hansen and Rubin’s *Listening Post* [2] aggregated and displayed fragments of public online text. *En Clair* operates at a more intimate scale — a single live conversation — and introduces an AI layer that actively transforms rather than represents content. It is grounded in Coelho and Labrune’s framework of the Large Language Object (LLO): AI understood as a design material embedded in physical form, inseparable from the artifact that carries it [1]. The NFC cartridge system draws on tangible computing traditions [3] in which physical tokens carry computational identity; here, they carry perceptual identity.

2 System Design

2.1 Hardware

The device is built around a **ESP32-S3** devkit microcontroller mounted on a custom double-sided FR1 PCB milled on a Bantam desktop mill. An **INMP441 MEMS microphone** (I2S) captures the speaker’s voice. A **Waveshare 2.13” e-ink display** (SPI) serves as the device’s status and cartridge-interaction interface. A **PN532 NFC transceiver module** (I2C) reads the NFC tags embedded in cartridges. Four physical buttons handle the full interaction: listen, cartridge-listen (for programming), confirm, and eject.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICMI '25, Canberra, Australia

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2.2 Software Pipeline

The ESP32-S3 communicates over USB serial with a **Node.js/Express server** running on a laptop. The pipeline has two modes:

Cartridge programming. When an unseen NFC tag is inserted, the user holds the cartridge-listen button and speaks a natural-language description of the perceptual filter they want — e.g., “someone who always assumes the worst,” or “brutally honest.” Audio streams to the server, is transcribed by Whisper, and then sent to Claude with a meta-prompt that generates a system prompt suitable for use as a perceptual filter. The system prompt and a short filter name are stored in a JSON database indexed by tag UID.

Filtered listening. During a conversation, pressing listen activates the microphone. The captured audio is transcribed, sent to Claude with the cartridge’s stored system prompt, and the filtered text is converted to speech by ElevenLabs and played through the headphones. Conversation history is maintained across turns to allow context-aware filtering.

2.3 Interaction and Physical Design

En Clair clips to the user’s waist. Opening the hinged enclosure to insert a cartridge simultaneously rotates the e-ink display upward via an abstracted scissor linkage, so the screen faces the user precisely when it is needed — during the cartridge moment — and folds away during listening. The filter name and summary appear on screen; the user confirms or re-records. Closing the device returns it to listening posture.

The e-ink display was chosen over higher-resolution or touch screens deliberately. Its slow refresh and visual restraint resist the register of consumer electronics, give the device an object-like quality, and constrain the screen to communicating only what is necessary during the cartridge interaction. The NFC cartridge system eliminated the need for complex electrical solutions to individuate cartridges: the tag carries the UID, the voice prompt carries the meaning.

The scissor linkage is the mechanical argument of the design. The screen’s orientation is physically coupled to the act of opening the device, encoding the logic spatially: the user does not navigate to a settings menu; they open a hinge.

3 Discussion

En Clair is ultimately a device about self-awareness. Its proposition is not that communication is broken, but that we all filter constantly and mostly without noticing — and that a named, embodied, audible filter might make those habits more legible.

Every relationship involves characteristic filtering. When speaking to a parent, many people instinctively soften bad news and manage the anticipated response. When raising an annoyance with a roommate, some are extremely direct while others are so indirect the concern never surfaces. A person who tends toward suspicion hears neutral feedback as criticism; a committed optimist hears the same as encouragement. *En Clair* offers a chance to encounter these patterns from the outside. Selecting a filter — “suspicious,” “projecting,” “optimist” — and hearing it applied to real speech in real time confronts the user with what that interpretive stance actually produces.

The cartridge system accommodates a wide range of filter types. Psychologically grounded filters model perceptual stances people actually inhabit. More playful ones — such as an 90s R&B love song lyrics — make the mechanics of transformation legible by exaggerating them: when “can you take out the trash” becomes a declaration of longing, the gap between input and output becomes the obvious subject of attention. Users who laughed at the R&B filter often engaged more carefully with the serious ones.

A more complicated finding emerged from filters defined using cultural archetypes. A filter described as “an Asian mom” consistently transformed expressions of praise into expressions of insufficiency — “I’m proud of you” became “you can do better.” This was not designed; it emerged from the LLM’s training data. The device did not create this pattern; it made it audible. Whether the model surfaces a real perceptual experience or reifies a caricature is a question the device cannot answer; it reveals that LLMs carry latent cultural models that become perceptual models when embedded in a device like this. Naming a filter is also an act of invoking whatever the model has learned to associate with that name.

The pipeline introduces several seconds of latency between speech and filtered playback. In testing, the filtered voice arriving after the conversation moved on created a delayed encounter with an alternative version of what had just been said — sometimes disorienting, sometimes revelatory. The gap may itself be meaningful: a beat in which both versions of an utterance briefly coexist.

4 Conclusion

En Clair makes the invisible work of perceptual filtering audible and tangible. By processing speech through a user-defined LLM filter encoded on a physical NFC cartridge, it produces a material encounter with the constructed nature of what we hear. The device does not improve communication; it reveals what is always already happening in it — and invites the listener to recognize, in the distortion, something of themselves.

Safe and Responsible Innovation Statement. *En Clair* is designed as a reflective art installation, not a covert surveillance or manipulation tool. All audio processing requires active user initiation and consent of the listener who wears the device. Filters that invoke cultural identities may reproduce stereotypes latent in LLM training data; this is surfaced as a finding rather than suppressed, and future work will examine how users from affected groups experience such outputs.

Acknowledgments

We thank Marcelo Coelho for guidance throughout the development of this project as part of MIT 4.043, Interaction Intelligence.

References

- [1] Marcelo Coelho and Jean-Baptiste Labrune. 2024. Large Language Objects: The Design of Physical AI and Generative Experiences. *Interactions* 31, 4 (July–August 2024), 43–48. <https://doi.org/10.1145/3672534>
- [2] Mark Hansen and Ben Rubin. 2001. Listening Post. Interactive installation. Brooklyn Academy of Music, New York.
- [3] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. In *Proceedings of CHI '97*. ACM, 234–241. <https://doi.org/10.1145/258549.258715>
- [4] Alexa Wright. 2000. *Conversation Piece*. Interactive installation.