

DIA 1 | 1 hora

Introduccion al Computo Afectivo

Emociones, modelo de Ekman, aplicaciones reales y el problema de la violencia familiar

¿Que son las emociones?

Las emociones son respuestas biologicas automaticas que nos ayudan a sobrevivir, comunicarnos y tomar decisiones.



Alegria

Placer,
satisfaccion



Tristeza

Perdida,
dolor



Enojo

Frustracion,
amenaza



Miedo

Peligro,
huida



Sorpresa

Evento
inesperado



Asco

Rechazo,
proteccion

Modelo de Ekman (1971)

Paul Ekman descubrio que existen 6 emociones basicas UNIVERSALES, reconocibles en todas las culturas del mundo, incluso en tribus aisladas de Papua Nueva Guinea. Se expresan en la cara, la voz y el cuerpo de manera involuntaria.

Nuestro proyecto: detectar el ENOJO en la voz porque precede a la agresion verbal y fisica.

¿Como expresa emociones la voz?

La voz no solo transmite palabras — tambien transmite emociones a traves de características acusticas.



Tono (pitch)

Frecuencia fundamental de la voz. Enojo: tono mas ALTO y variable. Tristeza: tono mas bajo y monotono.



Volumen (intensidad)

Que tan fuerte suena. Enojo: volumen ALTO. Miedo: volumen bajo. Sorpresa: cambio subito de volumen.



Velocidad (rate)

Que tan rapido hablas. Enojo: habla RAPIDA y entrecortada. Tristeza: habla lenta y pausada.



Timbre (calidad)

El 'color' de la voz. Enojo: voz ASPERA y tensa. Alegria: voz clara y brillante.

Los MFCCs capturan automaticamente estas características — por eso son la 'huella digital' del audio.

¿Que es el computo afectivo?

Definicion

El computo afectivo (Affective Computing) es el area de la IA que estudia como las maquinas pueden:

1. Reconocer emociones humanas
2. Interpretar su significado
3. Responder de forma apropiada

Termino acunado por Rosalind Picard del MIT en 1997 en su libro 'Affective Computing'.

Usa sensores como microfonos, camaras, pulseras y electroencefalografia.

Tipos de deteccion emocional



Facial

Camaras analizan expresiones (invasivo)



Fisiologica

Pulseras miden ritmo cardiaco, sudoracion



Texto

Analisis de sentimiento en mensajes



Voz (la nuestra)

Voz es el metodo MENOS invasivo porque no necesita camara ni contacto fisico.
Microfono analiza tono, volumen, timbre

¿Por que voz? Es no intrusiva, funciona con microfono barato (\$2 USD), no requiere que la persona se ponga nada, y funciona en la oscuridad (a diferencia de camaras).

Aplicaciones reales del computo afectivo

Salud

Oura Ring

Anillo inteligente que monitorea estres y calidad de sueno midiendo variabilidad cardiaca y temperatura corporal. Usado por la NBA para detectar COVID temprano.

Asistentes

Alexa / Siri / Google

Los asistentes de voz ya detectan frustracion en tu tono y adaptan sus respuestas. Amazon tiene patente para detectar emociones y sugerir productos.

Empresas

Call Centers

Empresas como Cogito y Beyond Verbal analizan llamadas en tiempo real para detectar clientes enojados y alertar al supervisor antes de que escale.

Seguridad

Automotriz

Mercedes y BMW integran camaras que detectan somnolencia y enojo del conductor para prevenir accidentes. Tesla monitorea atencion visual.

Entretenimiento

Videojuegos

Juegos adaptativos que miden frustracion del jugador y ajustan dificultad en tiempo real. Investigacion activa en MIT Media Lab.

El problema: violencia familiar en Mexico

70.1%

de mujeres han
experimentado violencia

INEGI, 2023

43.9%

violencia de pareja
a lo largo de la vida

ENDIREH, 2021

60%

mas prob. de baja salud
mental si abuso verbal infantil

Bellis et al., 2025

\$10

USD costo de nuestro
dispositivo detector

Este proyecto

¿Por que un dispositivo de \$10 USD?

Los sistemas actuales de deteccion emocional requieren computadoras (\$500+) o celulares (\$200+). Las familias mas afectadas por violencia son tambien las de menores recursos (CONEVAL, 2022). Un microcontrolador ESP32 de \$10 permite crear un sistema accesible que funciona sin internet, consume poca energía y puede operar 24/7 en el hogar.

Ref: Fernandez-Morales et al. (2025) IEEE Trans. Instrumentation and Measurement.

¿Como funciona nuestro sistema?

1

Captura

El microfono INMP441 captura 2 segundos de audio a 16,000 Hz via protocolo I2S digital. Resultado: 32,000 numeros.

2

Normalizacion

Se aplica pre-énfasis (resaltar agudos), filtro DC (quitar offset) y se detecta si hay voz real (VAD). Si es silencio, no procesa.

3

Extraccion MFCC

Se calcula la 'huella digital' del audio: ventaneo Hamming → FFT → filtros Mel → log → DCT. Resultado: matriz 20×63.

4

Cuantizacion

Los 1,260 valores float se convierten a int8 (enteros de 8 bits) usando: $q = \text{round}(\text{valor} / \text{escala}) + \text{punto_cero}$.

5

Inferencia CNN

MicroLightCNN procesa la matriz: 3 bloques Conv2D+BN+ReLU+Pool → Flatten → L2-Norm → Dense(64) → Sigmoid. Tiempo: ~219 ms.

6

Salida

Si probabilidad $\geq 70\%$: buzzer suena + LED muestra cara enojada. Si no: cara neutral. Dashboard en Serial Monitor.

Conociendo el ESP32-S3-N16R8

¿Que es un microcontrolador?

Una computadora completa del tamaño de un sello postal.
Tiene procesador, memoria, WiFi y Bluetooth.

A diferencia de tu computadora:

- Consume MUY poca energía (0.2W vs 65W)
- Cuesta ~\$10 USD (vs \$500+)
- No tiene pantalla ni teclado
- Pero ejecuta IA en tiempo real

N16R8 significa:

N16 = 16 MB Flash (no volátil, guarda el modelo)
R8 = 8 MB PSRAM (volátil, rápida, para cálculos)

El modelo MicroLightCNN pesa solo 22 KB — cabe 700 veces en la Flash.

Especificaciones técnicas

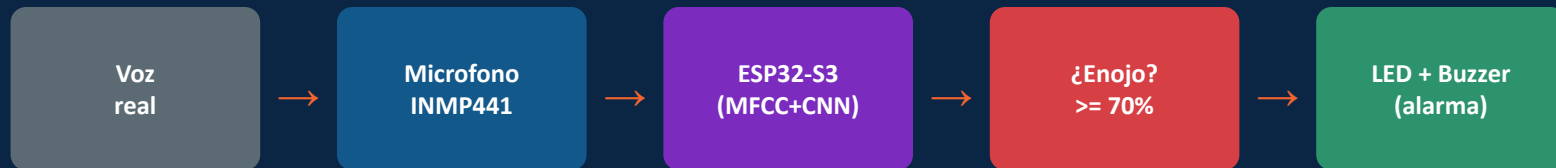
Procesador	Dual-core 240 MHz (Xtensa LX7)
Flash (N16)	16 MB — almacena programa + modelo
PSRAM (R8)	8 MB Octal — RAM de trabajo
WiFi	802.11 b/g/n (lo apagamos)
Bluetooth	BLE 5.0 (lo apagamos)
GPIO	45 pines programables

Periféricos que conectaremos:

Microfono INMP441 (I2S digital, \$3 USD)
Matriz LED 8x8 MAX7219 (SPI, \$2 USD)
Buzzer piezoeléctrico (GPIO, \$0.50 USD)
Total hardware: ~\$15 USD

Vision del proyecto final

Cada equipo construira un 'Jarvis' que escucha el enojo



¿Que van a aprender?

1. Como funciona el sonido digital
2. Procesamiento de audio (MFCCs)
3. Redes neuronales convolucionales
4. Programacion de microcontroladores
5. Integracion hardware + software + IA

Al final de la semana:

Presentan ante el grupo su sistema funcionando: hablan al microfono, el ESP32 procesa la voz en ~ 676 ms, y la matriz LED muestra si estan enojados o no.

Cada equipo tendra su propio modelo entrenado y su propio circuito armado.