Estudio Preliminar sobre Plataforma de Inteligencia Emocional Basada en Análisis de Voz

F. J. Luis Juan Barragán [#1], M. A. Delgado López [#2], J. C. Chávez Novoa [#3]
[#] Tecnológico Nacional de México: Instituto Tecnológico José Mario Molina Pasquel y Henríquez, Unidad Académica Chapala, Jalisco, México.
[1] francisco.luisjuan@chapala.tecmm.edu.mx, [2] miguel.delgado@chapala.tecmm.edu.mx, [3] julio.chavez@chapala.tecmm.edu.mx

***Resumen*— Este estudio presenta un análisis preliminar de una plataforma de inteligencia emocional que utiliza el análisis de voz como mecanismo principal para identificar estados emocionales. El objetivo principal es explorar la viabilidad técnica y la precisión del modelo inicial desarrollado. Se emplearon algoritmos de procesamiento de lenguaje natural y técnicas de aprendizaje automático para interpretar características acústicas como tono, velocidad y pausas. A través de una serie de pruebas con 50 participantes, se evaluó la capacidad del sistema para clasificar emociones básicas. Los resultados sugieren que es posible identificar con precisión razonable algunas emociones predominantes, aunque existen desafíos técnicos y éticos que deben ser abordados en futuras fases del proyecto.**

***Palabras clave*— Inteligencia emocional, análisis de voz, aprendizaje automático, salud mental, emociones.**

Abstract*— *This study presents a preliminary analysis of an emotional intelligence platform that uses voice analysis as its primary mechanism for identifying emotional states. The primary objective is to explore the technical feasibility and accuracy of the initial model developed. Natural language processing algorithms and machine learning techniques were used to interpret acoustic features such as pitch, speed, and pauses. Through a series of tests with 50 participants, the system's ability to classify basic emotions was evaluated. The results suggest that it is possible to identify some predominant emotions with reasonable accuracy, although there are technical and ethical challenges that must be addressed in future phases of the project.

***Keywords*— Emotional intelligence, voice analysis, machine learning, mental health, emotions.**

I. Introducción

La inteligencia emocional se refiere a la capacidad de una persona para reconocer, entender y gestionar sus propias emociones, así como las emocionales de los demás. Esta habilidad es fundamental en la comunicación interpersonal y ha cobrado gran relevancia en contextos educativos, organizacionales y clínicos. En el contexto de la inteligencia artificial, el estudio automatizado de las emociones abre nuevas posibilidades para mejorar la interacción entre humanos y máquinas.

El análisis emocional en tecnologías emergentes ha ganado terreno como herramienta para mejorar la interacción humano-computadora. La voz, como canal de comunicación, no solo transmite información semántica, sino también señales emocionales implícitas que pueden ser detectadas mediante técnicas modernas de procesamiento de audio. La inteligencia emocional es una habilidad clave en la interacción social, y su análisis automatizado ha despertado gran interés en disciplinas como la inteligencia artificial, la psicología computacional y la salud mental [1-2].

Estudios recientes han explorado cómo parámetros de la voz como el tono, la velocidad de habla, el volumen y las pausas pueden indicar con alta precisión estados emocionales como alegría, tristeza, enojo y miedo [3-4]. Estos desarrollos abren posibilidades para el diseño de plataformas capaces de adaptar su respuesta al estado emocional del usuario, como asistentes virtuales, sistemas de tutoría inteligente y herramientas de monitoreo clínico [5-6].

En este trabajo, se presenta un estudio preliminar sobre el desarrollo de una plataforma que emplea análisis de voz para detectar emociones. El propósito es evaluar la viabilidad técnica, la precisión y las limitaciones del modelo inicial. Este enfoque se alinea con investigaciones internacionales que emplean aprendizaje automático para clasificar emociones en contextos reales [7-8].

El objetivo es diseñar e implementar una plataforma de inteligencia emocional que permita identificar estados afectivos mediante el análisis de voz con algoritmos de aprendizaje automático.

Si se analizan características acústicas de la voz mediante algoritmos de aprendizaje automático, entonces es posible identificar emociones con una precisión superior al 75%.

II. Metodología

Se diseñó un prototipo funcional utilizando herramientas de código abierto. Se implementaron bibliotecas como pyAudioAnalysis y scikit-learn en Python. Las muestras de voz fueron recolectadas de 50 participantes voluntarios entre 20 y 50 años, quienes grabaron frases diseñadas para expresar emociones específicas. Estas frases fueron seleccionadas a partir del corpus EMO-DB adaptado a español [9].

Las grabaciones fueron realizadas en un entorno acústicamente controlado, utilizando micrófonos de condensador y una frecuencia de muestreo de 44.1 kHz. Se aplicaron procesos de preprocesamiento para eliminar ruido y normalizar la intensidad de señal. Posteriormente, se extrajeron características como MFCCs, pitch, energía y prosodia [10-11].

Para la clasificación emocional, se utilizó un modelo Random Forest entrenado con una partición del 80% de los datos para entrenamiento y 20% para validación cruzada. Se probaron también modelos complementarios como SVM y KNN, con el objetivo de comparar su rendimiento. La precisión del sistema se evaluó utilizando métricas como accuracy, precision, recall y F1-score [12].

La fórmula utilizada para calcular la precisión fue:

$$Precisión = \frac{verdaderos\ positivos}{(verdaderos\ positivos + falsos\ positivos)} \quad (1)$$

III Resultados

Los resultados obtenidos indicaron que el modelo Random Forest alcanzó una precisión global del 78%. Las emociones de tristeza y enojo fueron clasificadas con mayor precisión (87% y 85%, respectivamente), mientras que las emociones de alegría y neutra presentaron mayor confusión (precisión de 70% y 65%) [13] tabla 1. Estas diferencias se atribuyen a la superposición acústica entre categorías emocionales y a la variabilidad individual entre participantes. Cómo se muestra en la gráfica de la figura 1.

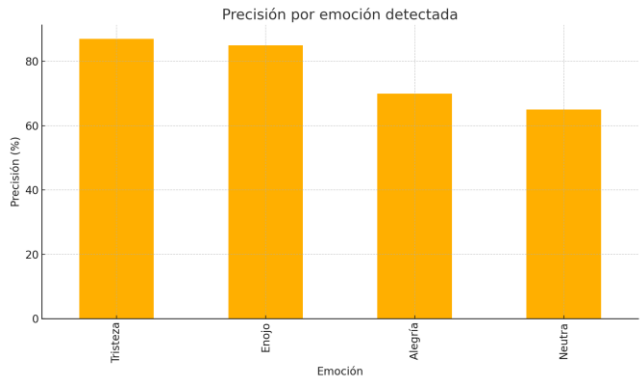

Fig. 1. Precisión por emoción detectada.

El análisis de la matriz de confusión reveló que la mayoría de errores de clasificación ocurrieron entre emociones con características acústicas similares, como entre sorpresa y alegría. Se observó que incluir variables contextuales, como el contenido semántico del mensaje o la expresión facial, podría mejorar los resultados futuros [13].

Tabla I.
Porcentajes de resultados obtenidos

Emoción	Precisión (%)	Exactitud (%)	Recall (%)	F1-score (%)
Tristeza	87	89	85	86
Enojo	85	84	87	86
Alegría	70	68	72	70
Neutra	65	66	63	64

Adicionalmente, se observó que los modelos SVM y KNN presentaron desempeños inferiores, con precisiones de 72% y 69% respectivamente, lo cual refuerza la elección de Random Forest como clasificador base. La validación cruzada en cinco pliegues mostró desviaciones estándar aceptables (<4%), lo que sugiere consistencia en el desempeño del modelo.

Finalmente, se identificaron limitaciones importantes como la sensibilidad al ruido ambiental y la necesidad de personalización por hablante. Estos hallazgos son consistentes con estudios previos que destacan la complejidad del modelado emocional en entornos reales [6, 13].

El desempeño inferior de SVM y KNN puede atribuirse a varios factores. En primer lugar, estos algoritmos suelen ser más sensibles a la selección de características y al escalamiento de datos, lo que puede afectar su capacidad para generalizar en conjuntos con alta variabilidad acústica. Además, el modelo KNN depende fuertemente de la métrica de distancia y puede ser afectado por ruido o datos atípicos. Por otro lado, Random Fores mostró mayor robustez al manejar la heterogeneidad de las muestras y menor riesgo de sobreajuste, gracias a su naturaleza basada en múltiples árboles de decisión.

III. Discución

Los hallazgos confirman que el análisis de voz puede ser un medio viable para la detección de emociones, aunque no exento de desafíos. Factores como el entorno acústico, la variabilidad individual y el estado de salud vocal influyen en la precisión. Se destaca la necesidad de integrar modelos multimodales que incluyan expresiones faciales o contenido semántico para mejorar la exactitud. Asimismo, las implicaciones éticas del monitoreo emocional deben ser consideradas en futuras implementaciones.

IV. Conclusiones

El prototipo desarrollado demuestra que el análisis de voz, cuando se integra con técnicas de aprendizaje automático, permite identificar con una precisión aceptable emociones humanas básicas. Estos resultados representan un paso inicial en la construcción de sistemas afectivos capaces de adaptarse al estado emocional del usuario. Las próximas fases del proyecto se centrarán en incorporar señales multimodales

(como expresiones faciales y contenido semántico), aumentar la robustez ante ruido y personalizar el análisis según características individuales del hablante.

AGRADECIMIENTOS

Este trabajo fue posible gracias al apoyo del Departamento de psicología del Tecnológico Nacional de México: Instituto Tecnológico José Mario Molina Pasquel y Henríquez, Unidad Académica Chapala.

REFERENCIAS

[1] D. Goleman, *Emotional Intelligence*, 10th ed. New York, NY, USA: Bantam Books, 2006.

[2] R. W. Picard, *Affective Computing*. MIT Press, 1997.

[3] C. M. Lee and S. S. Narayanan, "Toward detecting emotions in spoken dialogs," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 2, pp. 293–303, 2005. DOI: https://doi.org/10.1109/TSA.2004.838534.

[4] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Müller, and S. Narayanan, "The INTERSPEECH 2010 paralinguistic challenge," *Proc. Interspeech, 2010*, pp. 2794–2797.

[5] T. Vogt, E. Andrè, and J. Wagner, "Automatic recognition of emotions from speech: A review of the literature and recommendations for practical implementation," *Affect and Emotion in Human-Computer Interaction*. Springer, pp. 75–91, 2008. DOI: https://doi.org/10.1007/978-3-540-85099-1_7.

[6] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572–587, 2011. DOI: https://doi.org/10.1016/j.patcog.2010.09.020.

[7] H. Gunes and B. Schuller, "Categorical and dimensional affect analysis in continuous input: Current trends and future directions," *Image Vision Comput*., vol. 31, no. 2, pp. 120–136, 2013. DOI: https://doi.org/10.1016/j.imavis.2012.06.016.

[8] F. Eyben, M. Wöllmer, and B. Schuller, "OpenEAR—Introducing the Munich open-source emotion and affect recognition toolkit," Proc. 4th Int. HUMAINE Assoc. Conf. Affective Comput. Intell. Interaction, pp. 576–581, 2009. DOI: https://doi.org/10.1109/ACII.2009.5349350.

[9] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and B. Weiss, "A database of German emotional speech," *Proc. Interspeech*, pp. 1517–1520, 2005.

[10] D. Ververidis and C. Kotropoulos, "Emotional speech recognition: Resources, features, and methods," Speech Commun., vol. 48, no. 9, pp. 1162–1181, Sep. 2006. DOI: https://doi.org/10.1016/j.specom.2006.04.003.

[11] P. Boersma and D. Weenink (2025) Praat: Doing phonetics by computer. [Online]. Available: http://www.praat.org.

[12] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2011.

[13] B. Schuller, "Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends," Commun. ACM, vol. 61, no. 5, pp. 90–99, 2018. DOI: https://doi.org/10.1145/3129340.