

Identificación de reacciones emocionales mediante análisis cualitativo de audio y texto: un enfoque comparativo de metodologías

Laura Aguado-González, Yeray Sañudo, Jorge Sierra-Pérez

¹ Afiliación: Design for Safety (D4S)

Instituto de Investigación en Ingeniería de Aragón (I3A)

Universidad de Zaragoza, Mariano Esquillor s/n, 50018, Zaragoza, Spain.

Tel. +34-976762707, e-mail: laguado@unizar.es

Resumen

Los servicios oncológicos conllevan una gran carga emocional para el paciente y conocer sus reacciones durante el servicio permite introducir mejoras. Diferentes formas de análisis ayudan a detectar las emociones, ofreciendo diferentes resultados y reduciendo sesgos individuales. El análisis de audio es una opción preferible frente al análisis de texto.

Introducción

Los servicios oncológicos suelen conllevar una gran carga emocional que afecta a la toma de decisiones del paciente y repercute en los resultados del tratamiento [1]. En las primeras consultas, debido al miedo e incertidumbre, es donde se concentran más reacciones emocionales negativas. Conocer la experiencia de los pacientes y sus reacciones durante el servicio puede ayudar a introducir mejoras [2].

Las emociones pueden ser representadas por varios modelos, destacando las emociones discretas o espacios dimensionales [3], [4]. Los modelos dimensionales analizan las variaciones emocionales en función de 3 escalas: valencia, excitación y dominancia. Este modelo proporciona una comprensión y un análisis profundo, además de ser empleada en diversos campos por su fácil adaptabilidad [5].

Un análisis del discurso real durante las consultas puede aportar información sobre como responden los pacientes. Mediante métodos de análisis basados en la etnografía y técnicas de diseño de servicios, como el “shadowing” o el análisis de transcripción, puede comprenderse la experiencia del paciente [6], [7]. Aunque es un método sólido, hay que tener en cuenta algunas limitaciones [8]. Complementar este método con un análisis de audio puede ayudar a reducir sesgos y problemas de interpretación, ya que el audio proporciona información sobre el tono de voz, la velocidad, la intensidad o la variabilidad del habla de

los pacientes, lo que hace mucho más fácil y preciso detectar las reacciones emocionales. La mayor parte de la literatura encontrada sobre reconocimiento emocional a partir de audio trata de analizar el audio mediante algoritmos de aprendizaje automático y realizar esta tarea de forma automática.

Por este motivo, este estudio propone una forma manual de análisis de audio para la detección de reacciones emocionales a través de las dimensiones de valencia y excitación. Además, se realiza una evaluación de la calidad de los resultados obtenidos tanto por análisis de texto como de audio.

Metodología

Los datos son tomados durante las primeras consultas de oncología, observando las reacciones emocionales de los pacientes y grabando el audio. Posteriormente, el audio es transcrito y analizado. Las reacciones emocionales se toman siguiendo el modelo dimensional de valencia y excitación sin incluir la dominancia por su dificultad de medición, así se facilita el consenso y el análisis de los datos [9].

El análisis de texto empieza por la transcripción de los audios. Las transcripciones se analizan con el programa de análisis cualitativo NVivo [10] por dos personas de forma independiente, siguiendo las mismas escalas de las observaciones. Las reacciones son marcadas en base al contenido del texto y posteriormente se revisan aquellas respuestas en las que ha habido grandes discrepancias.

El audio se analiza de forma manual por dos personas de manera independiente. Debido a que en el audio hay numerosas variables a tener en cuenta, se crea una métrica para que ambos analizadores tengan los mismos criterios. Basándonos en la literatura [11], [12], [13], [14] se extraen qué variables del audio y qué propiedades caracterizan cada emoción. Dado que en la literatura las reacciones emocionales se miden por emociones discretas, se convirtieron estas

emociones en dimensiones de valencia y excitación basándonos en los estudios de [9], [15], [16]. Se desarrollaron dos métricas que evalúan de forma separada la valencia (Tabla 1) y la excitación (Tabla 2), con dos variables en cada una para disminuir la complejidad. De esta forma, todos los analizadores tienen criterios claros con los que juzgar los audios, facilitando la tarea y eliminando posibles sesgos.

Tabla 1. Métrica de análisis para valencia.

	Muy negativo	Negativo	Neutro	Positivo	Muy positivo
Contenido de audio	Muy negativo	Negativo	Neutro	Positivo	Muy positivo
Intensidad	Bajo	Medio	Moderado	Bajo	Alto

Tabla 2. Métrica de análisis para arousal.

	Calmando	Menos calmado	Algo excitado	Excitado	Muy excitado
Velocidad	Lenta	Moderada	Media	Rápida	Muy rápida
Variabilidad	Baja	Baja	Moderada	Alta	Alta

La correlación entre analizadores se evalúa mediante la estadística Kappa tanto en análisis de texto como de audio. Además, se realiza una triangulación de los resultados del análisis con las observaciones para ver que forma de análisis proporciona mejores resultados.

Resultados

Se han analizado las reacciones emocionales de 10 pacientes en el servicio de oncología, haciendo un seguimiento desde la primera consulta hasta el tratamiento, contando con la participación de tres oncólogos y dos enfermeras.

Las observaciones durante las consultas recogieron un total de 169 reacciones emocionales, mientras que el análisis de texto y audio solo detectó 86. Esto se debe a que muchas de las reacciones observadas no contenían audio del paciente, simplemente eran reacciones faciales o corporales.

La tabla 3 muestra los resultados de la estadística Kappa cuadrática ponderada, tanto de la correlación entre analizadores como entre tipos de análisis. La correlación entre analizadores en la transcripción es aceptable, mientras que la correlación entre analizadores en el audio es correcta, tanto para valencia como excitación. Esta mejora puede ser consecuencia de tener un mayor contexto y también de las métricas propuestas que proporcionaban un

criterio claro para ambos analizadores. El índice Kappa entre las observaciones y la transcripción no es suficiente para asegurar la correlación. Sin embargo, esta correlación aumenta considerablemente entre las observaciones y el análisis de audio. Esto se debe a que el audio incluye más información, como la intensidad o la velocidad de la voz, que ayudó a los analizadores a comprender mejor las emociones expresadas por los pacientes.

Tabla 3. Kappa de Cohen (cuadrática ponderada).

	Transcripción	Observación-transcripción	Audio	Observación-audio
Valencia	0.713	0.681	0.824	0.857
Excitación	-	-	0.862	0.620

Conclusiones

El uso de diferentes métodos de medición emocional puede proporcionar una visión más amplia de las emociones de los pacientes y reducir sesgos individuales.

El análisis de transcripciones no es capaz de detectar la dimensión de excitación, ya que al solo disponer de texto algunos aspectos como el tono o la intensidad no son posibles de identificar. Otros matices del discurso también se pierden, por lo que los resultados de valencia a veces pueden ser engañosos o difíciles de disitnguir. La correlación de resultados puede ser mejorable, además de que la imposibilidad de detectar la dimensión de excitación es algo alarmante. Por este motivo, se propone analizar también el audio de las consultas.

El análisis de audio proporciona unos resultados más robustos que las transcripciones, ya que aporta mayor contexto. La correlación de resultados fue aceptable. El desarrollo de criterios claros y sencillos fue clave para guiar a los analizadores. A pesar de que el análisis de audio manual es lento y requiere de más recursos, es una opción preferible frente al análisis de texto, debido a que aporta resultados de mayor calidad y permite detectar tanto valencia como excitación.

En futuros estudios, el análisis sentimental de audio basado en algoritmos podría reducir el tiempo de análisis; sin embargo, no hemos encontrado un algoritmo entrenado en este tipo de contexto, por lo que desarrollar uno específicamente entrenado para este entorno podría ser crucial para reducir tiempos y obtener resultados y conclusiones más rápidamente.

REFERENCIAS

- [1] C. Doyle, L. Lennox, and D. Bell, "A systematic review of evidence on the links between patient experience and clinical safety and effectiveness," *BMJ Open*, vol. 3, no. 1, p. e001570, Jan. 2013, doi: 10.1136/bmjopen-2012-001570.
- [2] K. Kafetsios, K. Hantzara, F. Anagnostopoulos, and D. Niakas, "Doctors' Attachment Orientations, Emotion Regulation Strategies, and Patient Satisfaction: A Multilevel Analysis," *Health Commun*, vol. 31, no. 6, pp. 772–777, Jun. 2016, doi: 10.1080/10410236.2014.993497.
- [3] M. E. Spinelli S, *Emotional Responses to Products. Methods in Consumer Research*, vol. 1. Elsevier, 2018.
- [4] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, "Text-based emotion detection: Advances, challenges, and opportunities," *Engineering Reports*, vol. 2, no. 7, Jul. 2020, doi: 10.1002/eng2.12189.
- [5] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J Behav Ther Exp Psychiatry*, vol. 25, no. 1, pp. 49–59, Mar. 1994, doi: 10.1016/0005-7916(94)90063-9.
- [6] Y. Sañudo, C. Akoglu, J. A. C. Rietjens, D. Snelders, A. M. Stiggelbout, and J. Sierra-Pérez, "The implementation of design methodologies for supporting shared decision making in healthcare services: A systematic review.," *Patient Educ Couns*, vol. 131, p. 108551, Feb. 2025, doi: 10.1016/j.pec.2024.108551.
- [7] A. S. Gallan, B. Perlow, R. Shah, and J. Gravdal, "The impact of patient shadowing on service design: Insights from a family medicine clinic," *Patient Exp J*, vol. 8, no. 1, pp. 88–98, Apr. 2021, doi: 10.35680/2372-0247.1449.
- [8] S. Point and Y. Baruch, "(Re)thinking transcription strategies: Current challenges and future research directions," *Scandinavian Journal of Management*, vol. 39, no. 2, p. 101272, Jun. 2023, doi: 10.1016/j.scaman.2023.101272.
- [9] K. R. Scherer, T. Baenziger, E. Roesch, U. Scherer, and V. Tran, "What are emotions? And how can they be measured?," *SAGE Publications*, vol. 44, pp. 695–729, 2005, doi: 10.1177/0539018405058216.
- [10] "Nvivo." Accessed: May 07, 2025. [Online]. Available: <https://www.nvivo-spain.com/>
- [11] A. Christy, · S Vaithyasubramanian, · A Jesudoss, and · M D Anto Praveena, "Multimodal speech emotion recognition and classification using convolutional neural network techniques," *Int J Speech Technol*, vol. 23, no. 3, pp. 381–388, 2020, doi: 10.1007/s10772-020-09713-y.
- [12] E. Oh, J. Suhr, and J. J. Lee, "Interplay of Lexical and Vocal Cues in Emotional Voice Perception," *International Journal of Advanced Culture Technology*, vol. 12, pp. 451–465, 2024, doi: 10.17703/IJACT.2024.12.4.451.
- [13] E. Preti, C. Suttora, and J. Richetin, "Can you hear what I feel? A validated prosodic set of angry, happy, and neutral Italian pseudowords", doi: 10.3758/s13428-015-0570-7.
- [14] K. A. Kumar and · J L Mazher Iqbal, "Machine learning technique-based emotion classification using speech signals," *Soft comput*, vol. 27, doi: 10.1007/s00500-023-08185-x.
- [15] Rohan Rajak and Rajib Mall, "Emotion recognition from audio,dimensional and discrete categorization using CNNs," IEEE, 2019.
- [16] B. Pan, K. Hirota, Z. Jia, and Y. Dai, "A review of multimodal emotion recognition from datasets, preprocessing, features, and fusion methods," *Neurocomputing*, vol. 561, p. 126866, Dec. 2023, doi: 10.1016/J.NEUCOM.2023.126866.