Chapter 1

Introducción

1.1 Motivación

Es indudable el impacto que ha creado la inteligencia artificial en la forma en la que nos comunicamos con las máquinas a día de hoy. La importancia de la interacción con las máquinas a través de comandos de voz, se ha visto acentuada gracias a la aparición de asistentes inteligentes como Siri (Apple) o Alexa (Amazon), que han explotado las diferentes áreas del análisis de la voz con el objetivo de mejorar la experiencia de usuario. Otras compañías como OTO han desarrollado modelos de análisis de voz capaces de detectar atributos únicos en la voz del interlocutor, lo que es usando por centros de asistencia telefónica para potenciar y mejorar sus sistemas automáticos. En definitiva, desde el primer software de reconocimiento por voz que fue presentado por IBM en 1961 reconociendo 16 palabras y dígitos, hasta la aparición de Google Home en 2017, ente tipo de asistentes han ido mejorando su alcance y capacidades.

El uso de estos asistentes no sólo se limita al ámbito doméstico, por ejemplo, la industria del videojuego creció un 23% durante la pandemia del Covid-19 en 2020 más que en el año anterior, 2019. La tendencia en el uso de este dominio empieza a extenderse hasta en la aplicación de asistentes personalizados para coches automáticos y asistencia de ayuda telefónica. Sin embargo, a pesar de los avances tecnológicos, estos asistentes de voz normalmente carecen de la habilidad de reconocer el estado emocional del usuario, y cerrar esta brecha podría ser un gran avance en las industrias ya mencionadas. Por ejemplo, Facebook usa inteligencia emocional para monitorizar signos de depresión en los usuarios.

Cabe pensar que en este tipo de tecnología haya un potencial interés para la asistencia sanitaria, o incluso, para la industria automovilística. Visualicemos por ejemplo, un conductor tratando de resolver una incidencia mientras conduce. Esta incidencia puede variar desde buscar una ruta al-

ternativa a un hospital o servicio de emergencia, y el estado emocional en el que se encuentre, puede afectar limitando su habilidad para resolver el problema.

De la misma manera el reconocimiento de emociones puede ocupar un lugar en los asistentes virtuales de cualquier servicio al integrarlo con técnicas del procesamiento del lenguaje natural, permitiendo mayor eficiencia del procesamiento de la conversación al detectar -por ejemplo- irritabilidad o frustración en el usuario.

El espectro emocional que una persona esconde en su discurso es un factor esencial de la comunicación humana y ofrece información sin necesidad de alterar el contenido lingüístico. Es aquí donde cabe preguntarse si ese reconocimiento emocional a través de la voz, está fuertemente ligado al idioma y la cultura, o hay emociones que podemos detectar independientemente de este. Por ejemplo, hay áreas con una diversidad cultura y lingüística muy diversa, sólo en Zimbabwe hay 16 lenguas oficiales, o 4 en Suiza. Yace aquí la necesidad de desvincular esta dependencia, lo que podría crear un impulso en el desarrollo de estos sistemas sin la necesidad de un corpus específico, y al mismo tiempo ayudarnos a entender la relación entre la expresión de emociones y la lengua. La idea de este proyecto nació por la motivación de crear un sistema capaz de crear una respuesta, no sólo coherente en el plano semántico, si no también sensible al estado emocional del usuario. Dado el alcance ambicioso con el que se partía, y teniendo en cuenta lo anterior, hemos preferido centrarnos en el estudio del reconocimiento de esas emociones, aplicándolo a la lengua extranjera.

1.2 Planteamiento del Trabajo

Desde hace años, el reconocimiento de emociones a través de la voz ha sido motivo de interés para la investigación, sin embargo siempre se ha estudiado sobre un mismo lenguaje debatiendo la habilidad de reconocer y clasificar las emociones oralmente expresadas. Esta habilidad ha sido respaldada por numerosos artículos donde se concluye que es posible distinguir e identificar entre al menos cuatro emociones básicas (felicidad, tristeza, y enfado) a través de la voz (sin necesidad del procesamiento del lenguaje natural y por lo tanto de un contexto).

Atendiendo al estudio de las emociones expresadas según la lengua existen estudios donde se demuestra que individuos de diferentes culturas pueden reconocer emociones básicas en diferentes niveles, pero es menos abundante la evidencia de un acuerdo en cómo las emociones básicas son reconocidas desde la expresión vocal de un interlocutor.Análogamente el debate del reconocimiento de emociones en un plano intercultural también se ha enfocado a través del estudio de los gestos faciales en conjunto con la expresión vocal, donde se concluye los factores sociales tienen un gran impacto, ya que la

identificación de las emociones es más fácil para los miembros de la misma cultura que para los de otra distinta [7] y [6]. A pesar de ello hay una gran carencia de comparativas con respecto a la voz donde se demuestre una sólida influencia cultural, sin embargo parece claro que las dimensiones socio culturales que engloban nuestras interacciones pueden tener un gran impacto en nuestra comunicación dentro de un marco emocional.

Este trabajo de fin de máster se centra en el uso de técnicas basadas en redes neuronales para la clasificación de emociones en el tracto vocal en la lengua extranjera. Para acercarnos a este escenario, se parte del supuesto que dado un modelo entrenado en un lenguaje capaz de reconocer emociones en este, se evalúa en un idioma distinto que nunca ha formado parte del anterior conjunto de datos.

Con este estudio se pretende entender mejor la relación entre emociones e idioma y arrojar luz a preguntas como si hay emociones que sean, más fáciles de reconocer independientemente del lenguaje.

Bibliography

[1] CIGDEM BAKIR and MECIT YUZKAT. "Speech Emotion Classification and Recognition with different methods for Turkish Language". In: *Balkan Journal of Electrical and Computer Engineering* 6.2 (2018), pp. 54–60. ISSN: 2147-284X. DOI: 10.17694/bajece.419557.

[2] Si Chen, Yiqing Zhu, and Ratree Wayland. "Effects of stimulus duration and vowel quality in cross-linguistic categorical perception of pitch directions". In: *PLoS ONE* 12 (July 2017), e0180656. DOI: 10.1371/journal.pone.0180656.

[3] Assel Davletcharova et al. "Detection and Analysis of Emotion from Speech Signals". In: *Procedia Computer Science* 58 (2015), pp. 91–96. ISSN: 18770509. DOI: 10.1016/j.procs.2015.08.032. URL: http://dx.doi.org/10.1016/j.procs.2015.08.032.

[4] Marc D. Pell. "Influence of emotion and focus location on prosody in matched statements and questions". In: *The Journal of the Acoustical Society of America* 109.4 (2001), pp. 1668–1680. ISSN: 0001-4966. DOI: 10.1121/1.1352088.

[5] Marc D. Pell et al. "Emotional speech processing: Disentangling the effects of prosody and semantic cues". In: *Cognition and Emotion* 25.5 (2011), pp. 834–853. ISSN: 02699931. DOI: 10.1080/02699931.2010.516915.

[6] Marc D. Pell et al. "Factors in the recognition of vocally expressed emotions: A comparison of four languages". In: *Journal of Phonetics* 37.4 (2009), pp. 417–435. ISSN: 00954470. DOI: 10.1016/j.wocn.2009.07.005.

[7] Marc D. Pell et al. "Recognizing Emotions in a Foreign Language". In: *Journal of Nonverbal Behavior* 33.2 (2009), pp. 107–120. ISSN: 01915886. DOI: 10.1007/s10919-008-0065-7.

[8] Social Signal Processing. "Research paper Social Signal Processing !" In: (2015).