Chapter 1

Introducción

1.1 Motivación

El espectro emocional que una persona esconde en su discurso es un factor esencial de la comunicación humana y ofrece información adicional sin alterar el contenido lingüístico. Las tecnologías orientadas a convertir la voz en texto (*speech to text*) no tienen una forma segura de medir la calidad del diálogo de su interlocutor, impactando en negocios que hacen uso de estos avances (por ejemplo, centros de atención al cliente donde miden su grado de satisfacción).

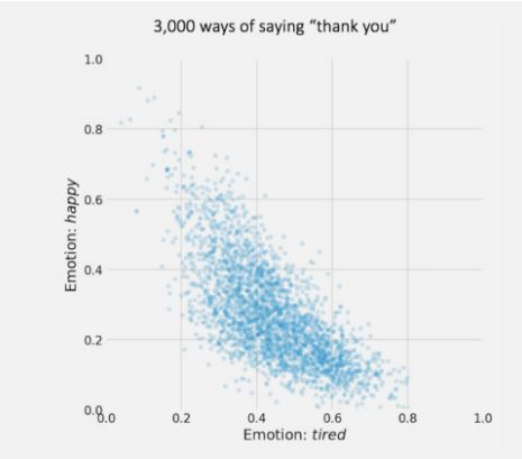

Figure 1.1: Dimensiones acústicas en la forma de decir "Thank you". Fuente: OTO

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)

[**Siri**] o Alexa (Amazon) [**amazonAlexa**] , 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 [**OTOPersonalAssistance**]. En definitiva, desde el primer software de reconocimiento por voz que fue presentado por IBM en 1961 reconociendo 16 palabras y dígitos[**IBMser**], hasta la aparición de Google Home en 2017 [**GHome2017**], este tipo de asistentes han ido mejorando su alcance y capacidades.

El uso de estos asistentes no sólo se limita al ámbito doméstico, ya que esta tendencia 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.

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 alternativa 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. SRI Ventures (una central estadounidense centrada en técnicas de procesamiento de la voz para el desarrollo de aplicaciones principalmente en el ámbito de la salud [**nuanceCom**]), desarrolla tecnología para analizar síntomas relacionados con enfermedades respiratorias, así como la aplicación de inteligencia artificial para evaluar por voz los sentimientos del cliente con el fin de mejorar el servicio.

Otro equipo de SRI liderado por Elizabeth Shriberg [**Shriberg2003**] decidieron combinar un sistema para entender la expresión oral con el análisis de emociones en la voz, dando como resultado una tecnología capaz de modelar computacionalmente la entonación de la voz del interlocutor derivando el significado sentimental más allá de las palabras usadas.

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 [**Pell2009a**] y [**Pell2009**]. 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 emoción e idioma y arrojar luz a preguntas como qué emociones son más fácilmente reconocibles indistintamente del lenguaje.

1.3 Estructura del trabajo

1. Introducimos de manera general, el contenido de las distintas partes que componen este trabajo. El primer capítulo introduce de manera esquemática el tema principal de este trabajo, justificando su importancia y el impacto en el mundo real. En este mismo capítulo se incluye:

    - La motivación, donde se argumenta la relevancia de este trabajo.

- El planteamiento del trabajo, que propone de manera general cómo solucionar el problema que vamos a encarar.

2. Seguidamente en el Contexto y estado del arte, se realizará una revisión de la literatura actual en relación con el tema a tratar. Se analizarán los resultados conseguidos hasta el momento así como las técnicas y métodos más usados en este ámbito, en concreto tendrá la siguiente estructura:

   - Se presentan las características fonéticas del habla y su trasfondo teórico, analizando las posibles limitaciones, o características a tener en cuenta a modo de introducción a las siguientes secciones, más técnicas.
   - El reconocimiento emocional del discurso propone una forma de modelar el problema haciendo uso de las técnicas que expone.
   - La extracción de características habla sobre qué características podemos encontrar en la voz para reconocer emociones y cómo podemos extraerlas.
   - Preprocesado de la señal describe el proceso para convertir la señal de audio a imagen de manera que se aprovechen mejor las ventajas de los clasificadores basados en redes convolucionales.
   - Los algoritmos de clasificación exponen diferentes métodos para categorizar las emociones una vez la señal está procesada.
   - La discusión sobre el estado del arte, revisa otros trabajos donde se han aplicado estas técnicas y cuáles han sido sus resultados.

3. La metodología de trabajo, comprende los objetivos generales y específicos que se han marcado para este trabajo así como el proceso que se seguirá para llevarlo a cabo.

4. El capítulo 4, plantea diferentes experimentos con el objetivo de llegar a una conclusión en la comparativa de esta tesis. Describe todos los componentes que formarán parte de esos experimentos de manera que se puedan reproducir siguiendo los pasos propuestos.

5. El capítulo 5, describe los resultados y el desarrollo de los experimentos planteados en el capítulo anterior, exponiéndolos de manera objetiva.

6. En el capítulo 6 se presenta una discusión de la conclusión sobre los resultados, comparándolos con otros trabajos y sus resultados.