

Detección automática de confrontación en una conversación entre interlocutores en archivos de audio.

Trabajo Terminal No. 2018-B127

Alumnos: Coxtinica Reyes Marco Antonio

Dirigidos: Dra. Jiménez Contreras Edith Adriana, Dr. Calvo Castro Francisco Hiram

Turno para la presentación del TT: TURNO Vespertino

e-mail: markescóm@outlook.com

Resumen – En este protocolo se propone el desarrollo de una aplicación para la detección automática de confrontación entre interlocutores en archivos de audio, que permitirá a los usuarios identificar a través del etiquetado de la señal, los puntos en el audio que sirvan como indicadores de la confrontación. En una primera etapa se hace el etiquetado en forma manual, y con los datos obtenidos, analizando las formas del audio, la aplicación procede a realizarlo automáticamente.

Palabras clave – Procesamiento de señales de audio, Tecnologías multimedia (herramientas de análisis de audio), Inteligencia Artificial, algoritmos para el reconocimiento de audio, y Psicología de la emoción.

1. Introducción

En la actualidad existen diversas aplicaciones para grabar y editar archivos de audio, detectar nombres de canciones, detectar sonidos cotidianos como alarmas contra incendios, el timbre de una casa y emiten vibraciones para las personas con discapacidad auditiva. También se han desarrollado aplicaciones que traducen audio a texto. Por lo que, el procesamiento de señales es un área de la Ingeniería Electrónica que ha permitido analizar archivos de audio, sonidos emitidos en la naturaleza entre otras situaciones, hallar patrones en las señales, encontrar características de interés, tener una aproximación del análisis del contenido de la información, lo cual, permite analizar, representar e interpretar el contenido de la información.

Existen aplicaciones que permiten realizar el etiquetado manual en las canciones, el cual consiste en el nombre de la canción, autor, año, duración entre otras características. Pero no permiten un etiquetado automático sobre la señal. Tales programas son: ID3 Tag Editor, Mp3 Tag, AHD ID3 Tag Editor, IDTE-ID3 Tag Editor y amp ID3 Tag Editor. Por lo que se busca utilizar un software para etiquetar señales de audio de forma manual.

Una etiqueta es una palabra clave generada por el usuario asociada con algún recurso, en nuestro caso audio[1].

Recientemente se ha prestado mucha atención a la predicción automática de etiquetas para música y audio en general. Ha habido muchos intentos de aplicar etiquetas automáticamente al audio para diferentes propósitos: gestión de bases de datos, recomendación musical, interfaces, estimado de similitud entre canciones, y así sucesivamente[1].

La detección automática de confrontación en una conversación entre interlocutores en archivos de audio, plantea aprovechar la representación de un archivo de audio como una señal, cuya gráfica nos permitirá identificar y etiquetar los diferentes picos que evidencien el estado que nos interesa detectar, en este caso el de confrontación, dentro de los posibles estados en las ondas de la señal, y con los datos obtenidos en una primera etapa de etiquetado manual, aprovechar esa información para entrenar a la aplicación para que pueda hacerlo de forma automática, mediante una herramienta como las wavelets.

Las wavelets conocidas como ondeletas u onditas son funciones que satisfacen ciertos requerimientos matemáticos y son utilizadas para la representación de datos o de otras funciones. Además, las wavelets son útiles para la aproximación de datos con variaciones o con discontinuidades abruptas [2].

Wavelets son herramientas para análisis de señal. El Wavelet Toolbox de MathLab está equipado con un conjunto de funciones y GUIs que permiten analizar señales e imágenes utilizando wavelets. Algunas funciones son: Compresión y eliminación de ruido en señales e imágenes, detección de puntos de ruptura y tendencias de señales, clasificación de señales y análisis multiseñal, transformadas wavelet continuas y discretas, análisis de paquetes wavelet y elevación de transformadas wavelet. [3]

En este trabajo, se requiere el análisis de archivos de audio para la detección de confrontaciones entre diversos actores, quienes participan en la confrontación. Entendiéndose por confrontación, el poner a dos personas en presencia una de otra, para comparar sus aserciones o para identificación entre sí; procesalmente, es el acto a través del cual se procura el reconocimiento que hace una persona respecto de otra que afirma conocer [4]. Otra definición utilizada es entendida como un choque relativamente violento entre fuerzas opuestas, y tiene como propósito reducir la ambigüedad e incongruencias en la experiencia y comunicación de la persona [5].

Se propone hacer uso de una herramienta de software matemático como matlab, para usar un algoritmo de clasificación de audio apoyado por el uso de etiquetas, que nos permitan clasificar los patrones de audio y las secciones donde se identifique el estado de confrontación en la conversación entre dos o más interlocutores.

28 Noviembre 2018
Vicente Corrales P.

Vo. Bo.
Ceciveline Arzola G.
29-Nov-18

Sistemas similares que se han desarrollado son.

1. Afectiva.
2. Hanson Robotics.
3. Amazon Alexa.
4. Google Home.
5. Siri

SOFTWARE	CARACTERÍSTICAS	PRECIO EN EL MERCADO
Afectiva	Análisis de Emociones a través de la voz, desarrollado por Afectiva.	Por confirmar.
Hanson Robotics	Hanson AI desarrolla una arquitectura cognitiva y herramientas basadas en AI que permiten a sus robots simular personalidades humanas, tener interacciones significativas con las personas y evolucionar a partir de esas interacciones. Operación puramente autónoma impulsada por la IA, está diseñada como un vehículo para la inteligencia híbrida de la IA humana. En modo híbrido, algunas de sus interacciones se generan de forma autónoma, y algunas están escritas por un equipo de desarrolladores de Hanson Robotics, o se navegan de forma semiautónoma con ayuda humana. Este modo de operación híbrido se usa comúnmente en telerobóticos y agentes de voz, y es una forma poderosa para que los humanos y las máquinas se ayuden entre sí. Esta complejidad tras bambalinas le permite crear conexiones emocionales y mantener conversaciones significativas con las personas. Estas interacciones le enseñan sobre lo que importa y valoran las personas. Este conocimiento invaluable le ayuda a continuar su camino hacia la verdadera autonomía y la sensibilidad . [6]	Según el modelo y capacidades del robot.
Amazon Alexa	Alexa es el servicio de voz ubicado en la nube de Amazon disponible en los dispositivos de Amazon y dispositivos terceros con Alexa integrada. Con Alexa, puedes crear experiencias de voz naturales para ofrecer a los clientes una forma más intuitiva de interactuar con la tecnología que usan a diario.	El precio del dispositivo incluye a la app Alexa.
Google Home	La app de Google Home permite configurar y controlar los dispositivos Chromecast (Chromecast, Chromecast Ultra, Chromecast Audio, TV o bocinas con Chromecast integrado).	El precio del dispositivo incluye a la app de Google Home.
Siri	Siri, asistente personal. Viene optimizado para tu iPhone, iPad, Mac, Apple Watch y Apple TV, y está siempre dispuesto a ayudarte. Además, Siri nunca deja de aprender. Por eso cuanto más lo uses, más útil te resultará.	El precio del dispositivo incluye a la app Siri.

Tabla 1. Resumen de productos similares.

2. Objetivo

Desarrollar una aplicación para la detección automática de confrontación entre interlocutores en archivos de audio para analizar, identificar y etiquetar en la señal el estado emocional que nos interesa detectar.

Objetivos Particulares

- Analizar un audio para marcar e identificar los puntos del estado de ánimo en la conversación entre interlocutores.
- Etiquetar en el audio, la parte de la señal que indica confrontación, de manera manual en una primera etapa, durante TT1.
- Etiquetar en el audio la parte de la señal que indica confrontación, automáticamente con los datos obtenidos del etiquetado manual, al aplicar un algoritmo de reconocimiento de audio, durante el TT2.

3. Justificación

La tecnología analiza la información a través del audio en un mensaje, pero la entonación también brinda información acerca del mensaje. El tono, el énfasis, la acentuación, la modulación de los interlocutores en la comunicación que establecen, resulta relevante para la interpretación, codificación y decodificación del mensaje.

La comunicación entre las personas cuenta con un componente emocional y afectivo que aporta información entre los interlocutores, información que puede pasar inadvertida, pero daría cabida a una mejor comprensión, análisis y resolución de conflictos.

La voz comienza a sustituir a los teclados como medio para ingresar información en los sistemas[7].

Detectar un estado emocional puede analizar el sentido en la toma de una decisión. Desde la decisión en la compra de un artículo de consumo, elegir a un candidato, un acuerdo académico, o mediar un tratado comercial.

Recientemente se ha prestado mucha atención a la predicción automática de etiquetas para música y audio en general. Es un esfuerzo por comprender mejor la tarea y también contribuir al aportar ideas para abordar este problema.

La creación de máquinas con sensibilidad emocional mejorará significativamente la interacción entre humanos y máquinas[8].

4. Productos o Resultados esperados

Interpretar los datos del estado de ánimo en información a través del etiquetado del audio, en una primera etapa se realiza el etiquetado de manera Manual, le enseñamos a la computadora a identificar los estados que nos interesan en la señal y en una segunda etapa se realiza el etiquetado Automáticamente apoyados en algoritmos ya existentes para clasificar audio. La arquitectura del producto a desarrollar se muestra en la Figura X.

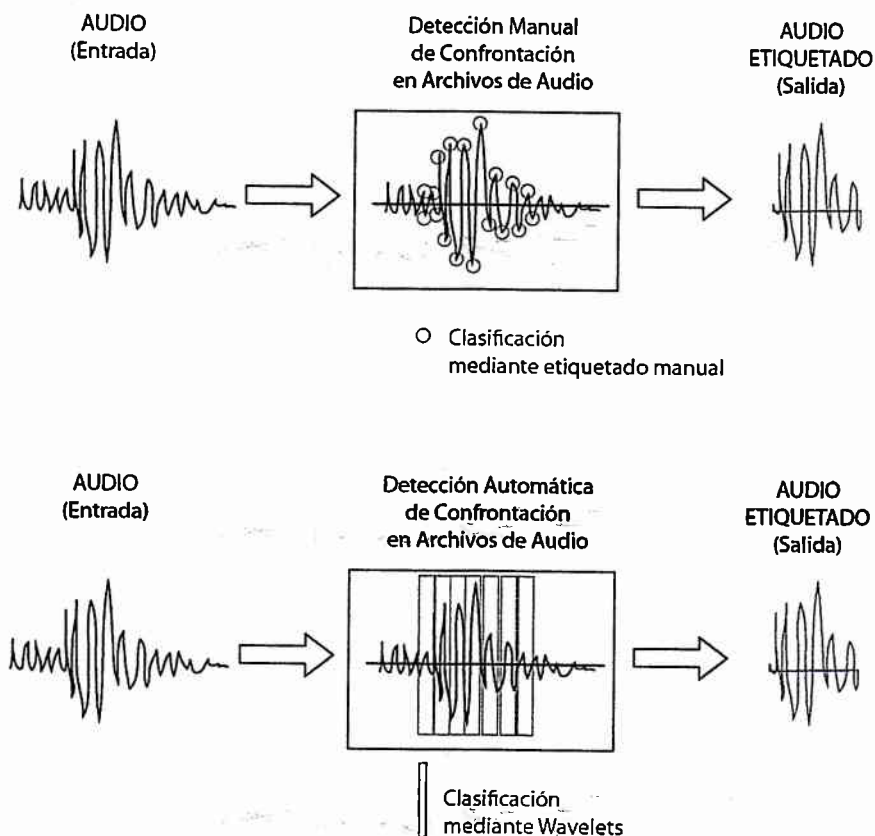


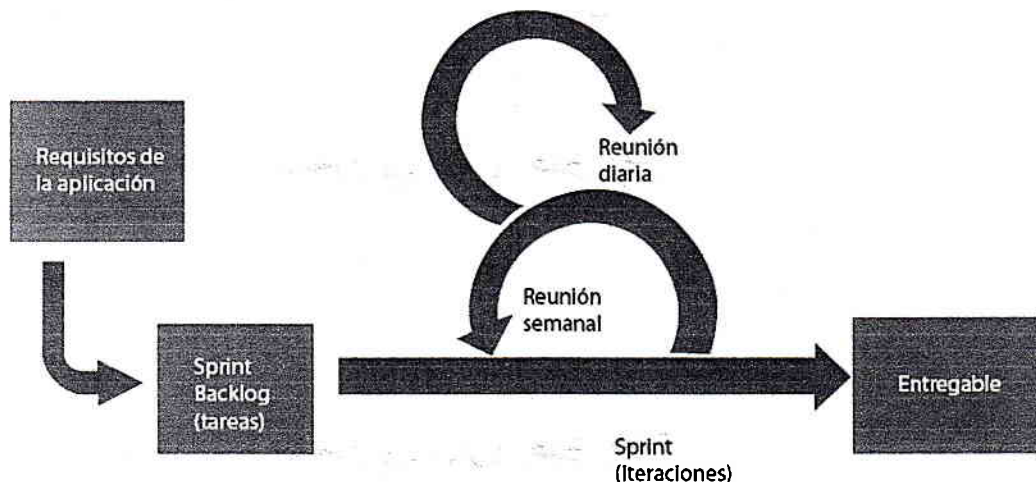
Figura X. Esquema de la aplicación.

Los productos esperados del TT, por ejemplo:

1. El código.
2. La documentación técnica de la aplicación.
3. El manual de usuario.

5. Metodología SCRUM

- **Product Owner:** Directores.
- **Equipo:** Los alumnos identifican y hacen las tareas asignadas.
- **Scrum Master:** Directores.
- **Ciclos (Sprints):** Por ejemplo en 16 semanas. Si tuviésemos 3 clases por semana, los Sprints podrían contener 36 clases de trabajo real (16 semanas x 3 clases/ semana menos las de hacer el backlog y las de review + retrospectiva). Un Sprint por semana, 36 sprints.
- **Stand-ups:** 5 minutos para actualizar el orden de tareas.
- **Diagrama de Burn-down:** Permite que el equipo vea cómo va su progreso en el Sprint, si va retrasado o avanzado.
- **Revisión:** Revisar los avances en el trabajo (presentación, documento).
- **Retrospectiva:** Revisar qué ha funcionado en ese Sprint / qué se tiene que mejorar.
- Volver a empezar el ciclo.



Se utilizará el aprendizaje automático mejor conocido como Machine Learning, el cual es una subdisciplina de la Inteligencia Artificial. Que ayudará en el reconocimiento de patrones, a través de archivos de audio en formatos WAV, MP3, MP4 y MIDI. Con la finalidad de detectar y simplificar la predicción de confrontación en una conversación entre interlocutores en un archivo de audio, en este caso conversaciones generadas en reuniones académicas o equipos de trabajo.

6. Cronograma

TT1.

Investigación de herramientas similares.

Investigación de algoritmos que permitan etiquetar el audio de manera manual.

Etiquetado de archivos de audio de forma manual tanto de las emociones como de los participantes que intervienen en el diálogo.

Actualmente, se cuenta con estos archivos de audio.

TT2.

Selección de herramienta y lenguaje de programación para el desarrollo de la aplicación que realizará de manera automática, el etiquetado automático donde se presente confrontación en los diálogos.

Realizar el procesamiento de análisis de señales.

Nombre del alumno(a): Coxtinica Reyes Marco Antonio

TT No.:

Título del TT: Detección Automática de Confrontación en Archivos de Audio

Actividad	ENE	FEB	MAR	ABR	MAY	AGO	SEP	OCT	NOV	DIC
Investigación bibliográfica										
Antecedentes de herramientas similares.										
Investigación y selección de herramienta que permite hacer etiquetado manual.										
Elaboración del etiquetado manual de archivos de audio (aproximadamente 10 archivos de al menos 1 hora de grabación).										
Pruebas iniciales										
Evaluación de TT I.										
Análisis diseño y desarrollo de una herramienta, que permitirá el etiquetado automática donde se detecten confrontaciones.										
Pruebas finales										
Elaboración del Manual de usuario e instalación.										
Generación el Reporte Técnico.										
Presentar los resultados en congresos.										
Evaluación de TT II.										


7. Referencias

- [1] Thierry BertinMahieux, Douglas Eck, y Michael Mandel. "Automatic Tagging of Audio: The State-of-the-Art". USA, Columbia University.
- [2] R. Castro, Liliana; Silva M., Castro. Wavelets y sus Aplicaciones. Depto. de Matemáticas. Depto. de Computación, Universidad Nacional del Sur. 1 er Congreso Argentino de Ciencias de la Computación. Bahía Blanca, Argentina.
- [3] <https://www.mathworks.com/videos/introduction-to-wavelet-toolbox-82002.html>
- [4] Vargas Cordero, Zoila Rosa, La confrontación: una oportunidad para el desarrollo personal. Revista Educación [en línea] 2003, 27 [Fecha de consulta: 20 de septiembre de 2018] Disponible en: <http://www.redalyc.org/articulo.oa?id=44027206> ISSN 0379-7082
- [5] Gómez Gutierrez, Emilia. "Representación de señales de audio". Escola Superior de Musica de Catalunya". <http://www.dtic.upf.edu/~egomez/teaching/sintesi/SPS1/Tema3-Representacion.pdf>
- [6] <https://www.hansonrobotics.com/sophia/>
- [7] Melgar, Jose'. "Análisis de emociones a través de la voz, desarrollado por Afectiva" <https://neuromarketing.la/2017/10/analisis-de-emociones-a-traves-de-la-voz/>
- [8] Cyril Lauriel, Perfecto Herrera. "Automatic Detection of Emotion in Music: Interaction with Emotionally Sensitive Machines". Barcelona, Universitat Pompeu Fabra.
- [9] <https://proyectosagiles.org/category/ejemplos/>

8. Alumnos y Directores

Marco Antonio Coxtinica Reyes.- Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta: 2010630254, email markescom@outlook.com.

CARÁCTER: Confidencial
FUNDAMENTO LEGAL: Art. 3, fracc. II, Art. 18, fracc. II y Art. 21, lineamiento 32, fracc. XVII de la L.F.T.A.I.P.G.
PARTES CONFIDENCIALES: No. de boleta y Teléfono.

Firma:  Marco Antonio Coxtinica Reyes

TURNO PARA LA PRESENTACIÓN DEL
TRABAJO TERMINAL:
VESPERTINO

Calvo Castro Francisco Hiram.- Dr. en Ciencias de la Computación en 2006, Nivel SNI: 1 (2015-2018). Áreas de Interés: Inteligencia Artificial y Cómputo Científico. Tel. email: hcalvo@cic.ipn.mx. hiramcalvo@gmail.com.

Firma: 

Edith Adriana Jiménez Contreras.- Licenciatura en Matemáticas Aplicadas y Computación, UNAM. Maestría y Doctorado en Ciencias de la Computación, CIC-IPN, Profesora del IPN en la ESCOM (Dpto. de Formación Básica) Áreas de Interés: Tecnología de Software, Sistemas de Información, Optimización, lógica matemática, algoritmos matemáticos. E-mail: ejjimenez@ipn.mx, edithajc@hotmail.com.

Firma: 