

ROSTROS QUE SUENAN

RECONOCIMIENTO FACIAL Y DETECCIÓN DE
PÚBLICO COMO HERRAMIENTAS DE CONTROL
PARA SISTEMAS DE SONIDO

Trabajo de final de grado

Javier Zóccoli

Carrera

**Licenciatura en
Artes Electrónicas**

Año

2021

ROSTROS QUE SUENAN

AGRADECIMIENTOS

A Dios, por darme la capacidad y el tiempo para escribir este trabajo.

A Caro y Emma, mi esposa e hija, por la paciencia y apoyo constante.

A mis tutores, por el tiempo invertido en leer detenidamente todas estas líneas y por sus consejos, que fueron de infinita ayuda.

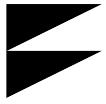
A Gabriela Golder, Lucía Kuschnir y Micaela Paz por haber aceptado la temática de mi trabajo, y por el acompañamiento constante durante todo el proceso.

Tutor

Dra. Cristina Voto
crivoto@gmail.com

Co-Tutor

Lic. Bernardo Piñero
bernardopinero@gmail.com



ÍNDICE

- 5** **Resumen**
- 6** **Abstract**
- 7** **Introducción**

- 11** **Capítulo 1**
- 11** Los inicios del reconocimiento facial
- 14** Biometría y los gestos
- 15** Técnicas aplicadas al reconocimiento facial
- 16** Detección de rostros
- 17** Sistema de reconocimiento facial
- 19** Técnicas y algoritmos
- 19** Eigenfaces
- 20** FisherFaces
- 20** Viola-Jones

- 22** **Capítulo 2**
- 22** Control del sonido
- 24** Acústica interior de la sala
- 25** La psicoacústica
- 27** Antecedentes

- 30** **Capítulo 3**
- 30** Proyección del dispositivo
- 31** Consideraciones en cuanto a la característica de las cámaras de video
- 33** Construcción del dispositivo - etapas ya materializadas
- 36** Construcción del dispositivo - etapas a completar
- 37** Reconocimiento facial con OpenCV y FaceRecognition
- 38** Creación de patch de audio en MAX
- 40** Consideraciones para la instancia práctica

42 **Conclusión**

43 **Bibliografía**

46 **Bío**

RESUMEN

El presente trabajo de investigación propone la elaboración de un sistema formado por un hardware y un software, capaz de brindar control y procesamiento de sonido utilizando los datos obtenidos mediante los algoritmos del reconocimiento facial y de la detección de público. En relación a la metodología empleada, es de tipo aplicada, ya que se propone efectuar aportes a una necesidad de carácter social (eventos, obras sonoras, salas, etc), en vista a aprovechar los recursos tecnológicos disponibles de reconocimiento facial en conjunto a la emisión y moderación de sonidos.

El entorno del sonido aún no encuentra límites en cuanto a la innovación de sus servicios por medio de la aplicación de nuevas tecnologías; es por ello que con el presente trabajo propongo la posibilidad de analizar la viabilidad y la futura expansión en cuanto al uso de esta técnica (reconocimiento facial), susceptible de generar y/o procesar contenido sonoro, permitiendo no solo la personalización de una escucha, según quien sea el espectador, sino también potenciar la experiencia de una audiencia o grupo de personas con sus respectivas particularidades. El proceso contribuye con múltiples áreas de investigación, como sistemas de seguimiento de gestos, modelado de la cantidad y cualidad de un público determinado, y la conexión entre el sonido y el gesto facial.

Palabras claves: **rostro, sistema de control, reconocimiento facial, audiencia, detección de público.**

ABSTRACT

This research work proposes the development of a system consisting of hardware and software, capable of providing sound control and processing using the data obtained through the algorithms of facial recognition and audience control. In relation to the methodology used, it is of an applied type, since it is proposed to make contributions to a need of a social nature (events, sound works, rooms, etc.), in view of taking advantage of the technological resources available for facial recognition in conjunction with the emission and moderation of sounds.

The sound environment still does not find limits in terms of the innovation of its services through the application of new technologies; That is why with this work I propose the possibility of analyzing the viability and future expansion in terms of the use of this technique (facial recognition), capable of generating and / or processing sound content, allowing not only the personalization of a listening, depending on who the viewer is, but also enhance the experience of an audience or group of people with their respective particularities. The process contributes to multiple areas of research, such as gesture tracking systems, modeling the quantity and quality of a given audience, and the connection between sound and facial gestures.

Keywords: **face, control system, facial recognition, audience, audience detection.**

INTRODUCCIÓN

El reconocimiento facial ha sido fuertemente estudiado durante los últimos años con diversas finalidades, mayormente destinadas al área de la seguridad. Sin embargo, se puede afirmar que han sido escasos los estudios que intentan vincular y obtener las ventajas que implicaría la combinación de estos dos campos, es decir la producción y procesamiento de sonido junto a una adecuada interpretación de las estructuras biométricas de los diferentes tipos de rostros y de las emociones que los rasgos faciales pueden expresar.

Por otra parte, aunque los humanos reconocemos caras familiares con poco esfuerzo, para automatización digitalizada esta tarea es enormemente compleja, pero asimismo, a medida que avanza la tecnología, la complejidad resulta cada vez menor. La causa principal de la enorme dificultad que entraña este problema, viene dada por las enormes variaciones existentes entre las distintas imágenes posibles de un mismo individuo. Son ejemplos de causas que producen dichas variaciones, los siguientes: la posición de la cara, su rotación respecto de las tres direcciones del espacio, la distancia entre la cara y el sensor, la escala, la resolución, el tipo de iluminación, el color, el maquillaje, la posición de la fuente de iluminación, la presencia de sombras, el peinado, la presencia de barba, bigote o perilla, y los gestos.

En este orden de ideas, resulta necesario detectar las necesidades en términos de algoritmos requeridos por los dispositivos, a los efectos de diseñar un sistema de sonido que sea capaz realizar un reconocimiento facial, logrando interpretar los datos y convirtiéndolo en un servicio de utilidad para el público definido.

La metodología utilizada es el análisis cuantitativo, cualitativo con la meta de analizar y explorar las necesidades en la elaboración de un sistema que sea capaz de brindar la creación, control y procesamiento de sonido mediante el reconocimiento facial.

Por otro lado, el trabajo es de intervención ya que, como lo afirma Stangnaro (2010), los proyectos de intervención se proponen la transformación de la realidad y se centran en el “hacer”. Más concretamente, se trata de la planificación de acciones articuladas que posibiliten que se debe intervenir en la realidad en vista de proporcionar al público alternativas de sistemas de

sonido con reconocimiento facial que ayuden a optimizar la experiencia del sonido en su más amplia gama.

Teniendo en cuenta lo anterior, se procederá a realizar un análisis de algoritmos de programación, para posteriormente efectuar pruebas con elementos básicos como cámara de video y audio; seguidamente, se implementarán herramientas de detección de rostros. Este paso, requerirá de la aplicación de técnicas de procesamiento de señales de audio que serán extraídos como información de los rostros observados. Todo esto se realizará en una interfaz sencilla en un comienzo para que así facilitar su uso a los usuarios que lo utilicen por primera vez e integrando los algoritmos más utilizados y eficientes que se conozcan, y que se hayan implementado en otros proyectos de la misma naturaleza. Finalmente, se logrará la integración completa del sistema.

Se debe considerar que este trabajo intenta dar una solución a las problemáticas que se observan en aquellos eventos o lugares en los cuales se desarrollan actividades que tengan que ver con lo sonoro, salas de concierto, iglesias, instalaciones sonoras, etc., y en los casos en donde este tipo de salas no cuente con una persona encargada o idónea para el control de todo lo que esté sonando, o que la extensión del espacio en cuestión exceda al control del operador de sonido operante.

Por otro lado, otro tipo de público al cual se encuentra dirigido este trabajo lo conforman grupos de personas con eventuales problemáticas especiales: personas de la tercera edad con molestias auditivas, personas con rasgos autistas (quienes presentan sensibilidad dentro de bandas de frecuencias altas), o quienes presenten dificultades para escuchar sonidos emitidos por parlantes y bafles, entre otros.

Además, se debe considerar la amplia versatilidad del producto final ya que podría ser aplicado a eventos masivos, obras sonoras en entornos poco explorados, conciertos con público itinerante, salas de presentación simultáneas, entre otros.

De este modo, se desprende la siguiente pregunta de investigación:

PREGUNTA DE INVESTIGACIÓN

¿Cuáles son las necesidades y requerimientos en torno a la creación de un dispositivo de reconocimiento facial con el cual se pueda controlar sistemas de sonido?

PREGUNTA

OBJETIVOS

Objetivo General

Analizar las necesidades y requerimientos en torno a la creación de un dispositivo de reconocimiento facial con el cual se pueda crear y/o controlar sistemas de sonido.

Objetivos Específicos

- a) Facilitar la percepción sonora a personas con eventuales problemáticas auditivas (personas de la tercera edad, personas con autismo)
- b) Reconocer la distribución de personas dentro de los límites de una sala, evento u obras sonoras.
- c) Adaptar y moderar la emisión de sonidos en espacios con escasa cantidad de público.
- d) Mejorar la inteligibilidad del sonido en recintos o salas de conferencia
- e) Identificar parámetros gestuales a nivel grupal en situación de eventos, con el propósito de brindar efectos sonoros acorde a la situación.

CAPÍTULO 1

En este apartado se pretende aproximarnos hacia el tema de reconocimiento facial, teniendo presente su evolución en los últimos años, las técnicas de reconocimiento facial, el proceso que requiere el mismo, con la idea de definir su importancia, y asimismo analizar sobre el estado de la cuestión.

Los inicios del reconocimiento facial

A algunos les puede parecer que el reconocimiento facial salió de la nada. Pero en verdad, esta tecnología ha estado trabajando durante algún tiempo. En este apartado compartimos un poco de la historia del reconocimiento facial para arrojar luz sobre cómo surgió esta tecnología transformadora y cómo ha evolucionado a lo largo del tiempo.

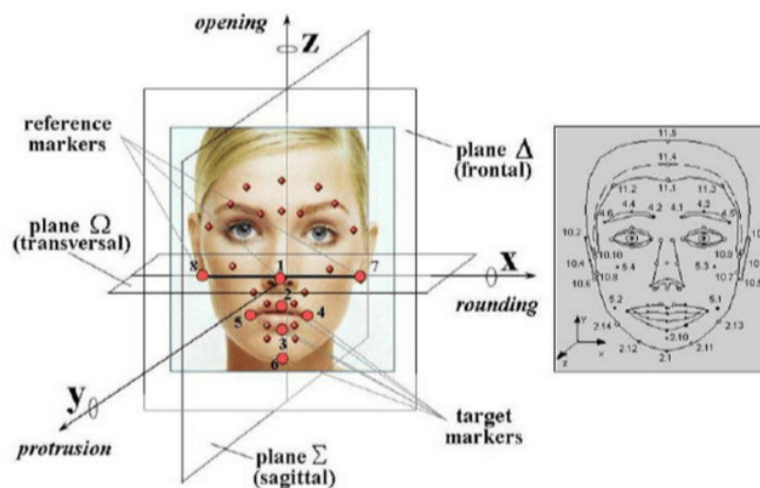
En los inicios de esta tecnología llamada “Reconocimiento facial” se usaba algoritmos de reconocimientos muy simples el cual daba mayor oportunidad a que los errores se produjeran, ya que al ser así el mismo reconocimiento se podría dar para 2 personas diferentes (Figueroa & Mellado, 2015). Sin embargo, en la actualidad y con los avances logrados, además de los algoritmos que han sido exponencialmente mejorados, los errores son mínimos ya que se han afinado la forma en cómo se reconoce cada rostro.

Con el advenimiento de la tecnología reciente se puede decir que el reconocimiento facial ha pasado por diversas etapas que van desde el campo de la ciencia ficción a ser temida como un dispositivo de control. No hay dudas que en la actualidad se encuentra presente en innumerables aplicaciones con finalidades distintas.

Si buscamos algún referente en la temática, se debe tener presente a Woodrow Wilson como el pionero en este tipo de tecnología; allá por los años 1960 había trabajado en un sistema para clasificar los rasgos del rostro humano por medio de una tabla RAND. Ello implicaba el uso de un lápiz óptico y unas coordenadas con la idea de situar los ojos, nariz o la boca de las personas de la manera más exacta posible, aunque resultaba un esquema aún muy manual (BeeDigital, 2019).

Estas métricas señaladas podrían entonces insertarse en una base de datos. Seguidamente, cuando el sistema recibía una nueva fotografía de un individuo, podía recuperar la imagen de la base de datos que más se parecía a ese individuo. En aquel momento, el reconocimiento facial estaba desafortunadamente limitado por la tecnología de la era y el poder de procesamiento de las computadoras. No obstante, fue un primer paso importante para demostrar que el reconocimiento facial era una biométrica viable.

Posteriormente, casi una década después llegarían Goldstein, y Lesk, que detallaron estas características faciales e iniciaron la mejora hacia la precisión del reconocimiento facial. En esta época pudieron agregar mayor precisión a un sistema de reconocimiento facial manual. Para ello usaron 21 marcadores subjetivos específicos que incluyen el grosor de los labios y el color del cabello para identificar las caras automáticamente. Al igual que con el sistema de Bledsoe, la biometría real aún tenía que ser computada manualmente (Spot, 2019).



Fuente: extraído de Spot, (2019) https://medium.com/@spot_blog/una-breve-historia-del-reconocimiento-facial-vision-blog-5a76fdfe4865

Más adelante, a finales de los años 80, se aplica el álgebra lineal, gracias a Sirovich y Kirby. En 1988 estos creadores, comenzaron a aplicar el álgebra lineal al problema del reconocimiento facial. Lo que se conoció como el enfoque Eigenface comenzó como una búsqueda de una representación de imágenes faciales de baja dimensión. En los avances que realizaron Sirovich y Kirby, lograron mostrar que el análisis de características en una colección de imágenes faciales podría formar un conjunto de características básicas. También pudieron demostrar que se requerían menos de cien valores para codificar con precisión una imagen facial normalizada.

Para inicios de 1991, Turk y Pentland elaboraron la tecnología idónea para detectar un rostro humano dentro de una fotografía, abriendo paso al reconocimiento facial automático.

Es así que, poco a poco el reconocimiento facial se fue haciendo un hueco en las aplicaciones de seguridad, en particular aquellas concernientes al Estado. En 2001, nace el Viola-Jones Object Detection Framework, que propone algoritmos para detectar objetos dentro de imágenes, y que enseguida fue utilizado para la detección de rostros de forma exitosa. En la década actual, aparecen los Convolutional Neural Networks, que hasta hoy suponen el mejor modo de detectar rostros, haciendo uso de sistemas muy potentes capaces de almacenar información en la nube (BeeDigital, 2019).

Durante el periodo que abarca 1993-2000, la Agencia de Proyectos de Investigación Avanzada de Defensa (DARPA) y el Instituto Nacional de Estándares y Tecnología lanzaron el programa de Reconocimiento Facial (FERET) con el fin de fomentar el mercado de reconocimiento facial comercial. El proyecto involucró la creación de una base de datos de imágenes faciales. La base de datos se actualizó en 2003 para incluir versiones en color de alta resolución de 24 bits. Incluido en el conjunto de prueba fueron 2.413 imágenes faciales todavía que representan 856 personas. La esperanza era que una gran base de datos de imágenes de prueba para el reconocimiento facial podría inspirar la innovación, que podría dar como resultado una tecnología de reconocimiento facial más poderosa (FaceFirst, 2018).

En cuanto a los medios masivos de comunicación, no fue hasta 2010 cuando esta tecnología llegó al público por medio de Facebook. Conforme el público general ha conocido esta tecnología, se ha convertido en una cuestión controvertida, pues muchos ciudadanos y medios de comunicación cuestionan la privacidad de sistemas de este tipo

(FaceFirst, 2018).

A pesar de los vertiginosos avances, en la actualidad no existe una aplicación que reconozca o identifique en una imagen a una persona, porque todavía es necesario el manejo de volúmenes de datos importantes. Para el software no entrenado, una imagen facial sólo es un puñado de píxeles que se convierten en valores numéricos en función de la intensidad y dirección de la luz y la sombra. A partir de aquí puede comenzar a identificar patrones que corresponden a rasgos faciales, cuencas oculares, narices, mandíbulas o mentones.

Por otro lado, son algunas grandes corporaciones como Google, Facebook, Apple, Amazon, entre muchas otras, también utilizan dicha tecnología para finalidades que van desde la seguridad hasta la mejora de la experiencia del usuario de un servicio en sus distintas variantes.

Durante el 2019 tuvo mucha popularidad la denominada FaceApp; es decir, la aplicación móvil protagonista del último viral de moda, consistente en subir nuestras fotos para obtener aproximaciones de cuál sería nuestro aspecto dentro de varias décadas (o siendo del sexo opuesto), está saltando a los titulares también por la amenaza que representa para nuestra privacidad (Suru, 2020).

Biometría y los gestos

La palabra biometría proviene del vocablo griego: bios que significa vida y metron cuyo significado es medición, por lo tanto la biometría es la ciencia que establece la identidad de un individuo a partir del análisis de las características fisiológica o del comportamiento del cuerpo humano (López, 2014).

En el terreno de la tecnología de la información, la biometría hace referencia a las tecnologías que miden y analizan las características del cuerpo humano, como el ADN, las huellas dactilares, la retina y el iris de los ojos, los patrones faciales o de la voz y las medidas de las manos a efectos de autenticación de identidades.

No caben dudas que la biometría se encuentra ligada al reconocimiento de los gestos. A su vez, esta última forma parte de las ciencias dentro del campo de la tecnología y del lenguaje y tiene como meta interpretar los gestos humanos por medio de algoritmos matemáticos.

Los gestos pueden ser cualquier movimiento corporal o estado, pero comúnmente se originan a partir de la cara o la mano. Enfoques actuales en el campo incluyen reconocimiento de la emoción facial y el reconocimiento de gestos de la mano. Muchos enfoques que se han hecho hacen uso de cámaras y visión (por medio de cámaras) y se procesan por medio de algoritmos para interpretar el lenguaje de señas. Sin embargo, la identificación y el reconocimiento de la postura, la marcha, la proxémica, y los comportamientos humanos es también el tema de técnicas de reconocimiento de gestos (Rehm, Bee, André, 2007).

La biometría facial de una persona cambia radicalmente para el sistema si el usuario decide llevar gafas o dejarse barba. Además, el rendimiento del sistema se ve muy afectado por cambios de iluminación y posición de la cámara y del usuario.

Técnicas aplicadas al reconocimiento facial

El avance del conocimiento en el área del reconocimiento facial se ha convertido en una de las aplicaciones más estudiadas dentro del campo como la biometría o el reconocimiento de patrones (Hernández, 2010). El interés sobre la temática tiene que ver con el incremento del número de soluciones comerciales, así como el crecimiento de este mercado; ejemplo de ello son los sistemas basados en el reconocimiento para la identificación criminal y sistemas de video vigilancia desarrollados por empresas informáticas.

A los efectos de determinar la identidad de una persona se puede tener en cuenta muchos aspectos; estos pueden ser tanto físicos, como pupilas, contorno de la cara, el iris, entre otros, o se puede tratar de aspectos de orden psicológicos como serían los gestos.

Uno de los métodos principales se denomina Análisis de Componentes Principales (ACP), y este transforma un número de variables posiblemente correlacionadas en un pequeño número de variables llamadas componentes principales, a través de un algoritmo de reducción que facilita encontrar los vectores que mejor representan la distribución de un grupo de imágenes.

Otro método interesante lo constituye el denominado transformación discreta del coseno: en síntesis, es una transformación que representa una secuencia finita de datos como la suma de una serie de funciones cosenoidales oscilando a diferentes frecuencias (Hernández, 2010). La señal

procesada se utiliza en aplicaciones de reconocimiento facial, obteniendo características de la imagen y luego se clasifican conforme a criterios de identificación.

Detección de rostros

La detección facial es una tecnología de visión computarizada que determina el lugar y el tamaño de rostros humanos en imágenes o videos. Es uno de los subtipos de detección de objeto, cuya tarea es encontrar la localización y el tamaño de los objetos en una imagen perteneciente a la clase dada (Wikipedia, 2020).

Mientras que para el ser humano es insignificante esta tarea, una computadora presenta dificultades debido a diversos factores como: variabilidad en la posición del rostro, presencia o ausencia de componentes estructurales (ejemplo: bigote), expresión facial, oclusión (uso de lentes, gorro, etc.) y condiciones del ambiente.

Detectar el rostro humano es el primer paso en un sistema de reconocimiento facial, sin embargo influye de manera significativa en el resultado del proceso, ya sea dado un conjunto de imágenes o video en tiempo real. Por lo tanto, debe ser capaz de identificar los rostros independientemente de los factores que anteriormente se mencionaron.

Según Espinoza & Jorquera (2015), los métodos de detección facial se dividen en cuatro categorías (no excluyentes):

- Métodos basados en conocimiento: Codifican el conocimiento humano mediante distancias y posiciones entre las características humanas (ojos, nariz, labios).
- Métodos basados en características invariantes: Las características invariantes son aquellas que no se modifican a eventuales cambios de luz, pose o ubicación de la cámara, tales como la ceja, nariz, textura de la piel y línea de pelo. Este método funciona detectando uno de estos componentes, construyendo un modelo estadístico y con los resultados, verificar la existencia de un rostro.
- Métodos basados en moldes (patrones): Es la relación entre una imagen de entrada y un patrón o molde previamente definido, cuyo objetivo es capturar características del rostro.

- Métodos basados en apariencia: Utilizan modelos obtenidos mediante entrenamiento de imágenes, tomando la imagen como un vector de características, es decir, es visto como una variable aleatoria. A diferencia de los métodos basados en moldes, donde el patrón es definido por un “experto”, los patrones en este modelo son determinados por el aprendizaje obtenido en el entrenamiento de imágenes (Espinoza & Jorquera, 2015, pág. 19).

Sistema de reconocimiento facial

El sistema de reconocimiento facial es una aplicación dirigida por una computadora que logra reconocer automáticamente a una persona en una imagen digital (Rearte & Sánchez, 2011).

El reconocimiento facial, en opinión de Rearte & Sánchez (2011), se ha convertido en los últimos años en un campo de estudio cada vez más activo que incluye diversas disciplinas vinculadas tanto a procesamiento de imágenes, reconocimiento de patrones, visión por ordenador y redes neuronales. Abarca tanto a investigadores del área de informática como a neurocientíficos y psicólogos.

Etapas para su funcionamiento

El proceso de reconocimiento facial presentado por Rearte y Sánchez consta de cuatro módulos principales:

1. Detección de la cara:

Detecta que hay una cara en la imagen, sin identificarla. Si se trata de un video, también podemos hacer un seguimiento de la cara. Proporciona la localización y la escala a la que encontramos la cara.

2. Alineación de la cara:

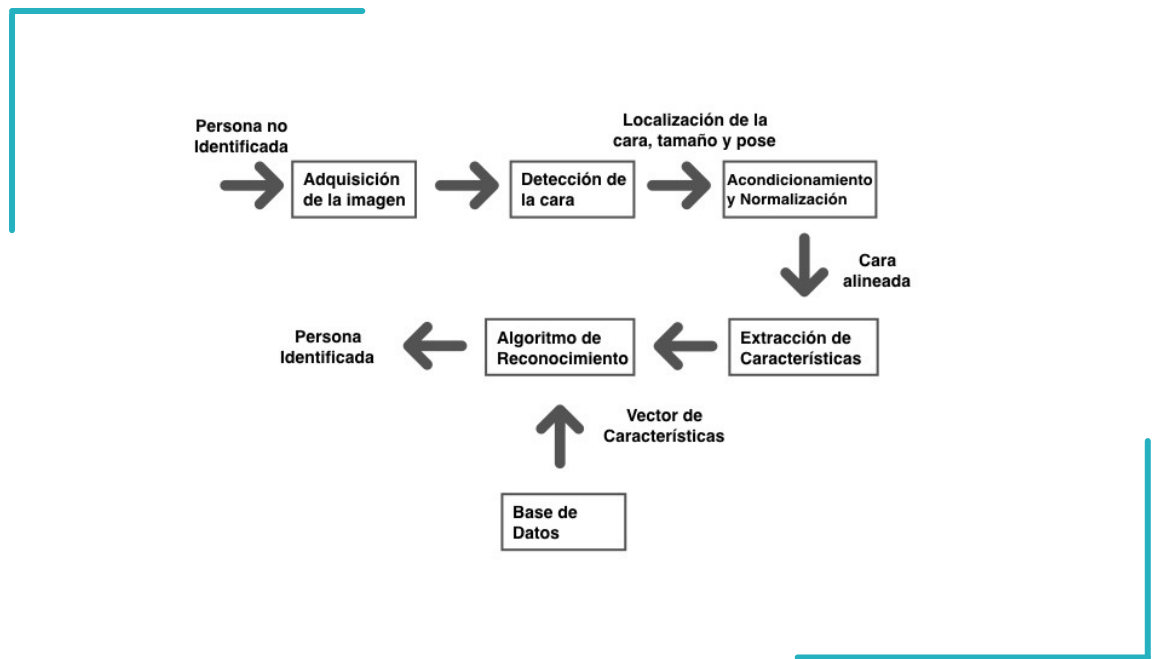
Localiza las componentes de la cara y, mediante transformaciones geométricas, la normaliza respecto propiedades geométricas, como el tamaño y la pose, y fotométricas, como la iluminación. Para normalizar las imágenes de caras, se pueden seguir diferentes reglas, como la distancia entre las pupilas, la posición de la nariz, o la distancia entre las comisuras de los labios. También se debe definir el tamaño de las imágenes y la gama de colores. Normalmente, para disminuir la carga computacional del sistema, se acostumbra a utilizar imágenes pequeñas en escala de grises. A veces también se realiza una ecualización del histograma.

3. Extracción de características:

Proporciona información para distinguir entre las caras de diferentes personas según variaciones geométricas o fotométricas.

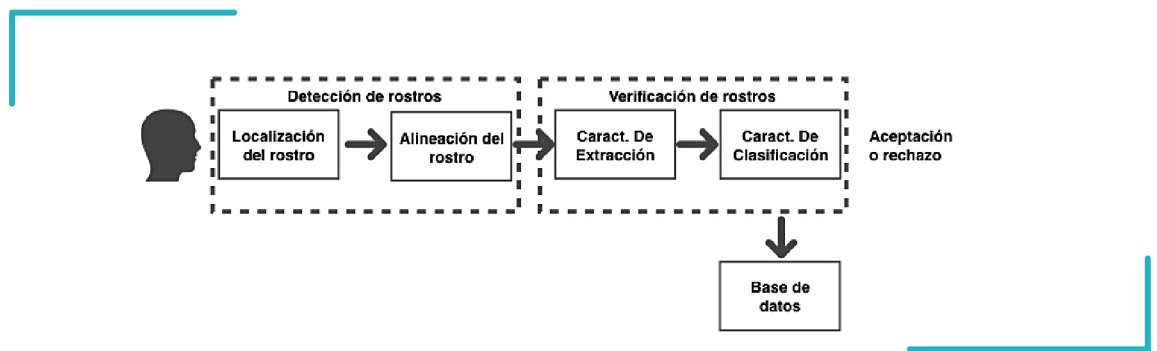
4. Reconocimiento:

El vector de características extraído se compara con los vectores de características extraídos de las caras de la base de datos o algoritmos. Si encuentra uno con un porcentaje elevado de similitud, nos devuelve la identidad de la cara; si no, nos indica que es una cara desconocida.



Fuente: recuperado de Rearte, Manuela, Sánchez Martín (2011). *Introducción a la Inteligencia Artificial*. Universidad Nacional de Rosario, Argentina.

En el modelo diseñado por Espinoza & Jorquera (2015), explica cómo funciona cada etapa dentro del reconocimiento facial, la cual consta de varias secciones en complemento.



Fuente: Esquema proporcionado por Figueroa & Mellado (2015).

El reconocimiento facial no puede realizarse sin las técnicas y algoritmos. En suma, son los propios algoritmos (base de datos disponibles) los cuales hacen posible la comparación de los rostros detectados en tiempo real.

Cabe señalar, que el campo de la definición de técnicas y algoritmos más adecuados pertenece a la disciplina de la ingeniería y programación, el cual estudian los lenguajes de programación y las herramientas necesarias para el reconocimiento facial. De este modo, se presenta un breve recorrido sobre técnicas y algoritmos que podrían ser más adecuados para el desarrollo del presente proyecto.

Técnicas y algoritmos

Como se ha mencionado en párrafos anteriores, la detección facial es la localización de las posibles caras que se pueden encontrar en una imagen, extrayendo para ello el centro de la cara, su orientación y su escala. Una vez detectada la cara se aplican algoritmos de seguimiento y/o reconocimiento, según el contexto y la finalidad de la aplicación.

A continuación se presenta algunos de los algoritmos desarrollados para la detección de diferentes tipos de rostros y reconocimiento facial, por lo tanto se va mostrar el uso de algoritmos ya testeado que haya demostrados obtener buenos resultados.

Eigenfaces

Por un lado se encuentra Eigenfaces, el cual es un algoritmo desarrollado por Turk y Pentland. Este método de reconocimiento facial se basa en el análisis por componentes principales. Identifica patrones en las imágenes y clasifica a cada individuo en función de las coordenadas obtenidas en el subespacio que se forma a través de estas componentes (Nerea, 2017).

En este sentido, con la finalidad de conseguir esta clasificación previamente, se necesita de un set de entrenamiento, es decir, se realiza una colección de diferentes imágenes de caras de personas que harán papel de vectores de observación. Estos vectores vendrán determinados por cada pixel de la imagen.

Para poder utilizar este algoritmo, todas las imágenes han de tener el mismo tamaño. Todas las imágenes deben tener $n = W \times H$ píxeles, donde W es el ancho de la imagen y H la altura.

Por lo tanto, cada imagen es representada como un punto en un subespacio vectorial de n dimensiones. Los desarrolladores de este algoritmo utilizaron imágenes de 256×256 píxeles, por lo que las imágenes estaban contenidas en un subespacio vectorial de 65536 dimensiones (Nerea, 2017).

FisherFaces

Esta técnica considera las imágenes de entrenamiento de un mismo individuo como clases, por lo tanto existen el mismo número de clases que personas. Una vez definida las clases se procede a calcular dos matrices: la matriz de dispersión entre clases y la matriz de dispersión dentro de clases. Una vez calculada estas matrices se obtiene una matriz de proyección donde cada columna será la base del nuevo sub-espacio, denominada Fisherfaces (1).

Viola-Jones

Viola-Jones es un algoritmo capaz de detectar objetos y caras en tiempo real. Este algoritmo fue propuesto por Paul Viola y Michael Jones en 2001. La característica principal que consigue hacer que este algoritmo sea uno de los más usados es su nivel de efectividad, ya que tiene una tasa de detección alta, tasa de verdaderos positivos muy alta y tasa de falsos positivos muy baja. Además, la utilización de este algoritmo supone un ahorro de tiempo considerable ya que no serán procesadas subregiones de la imagen que no se sepa con certeza que contienen una cara y sólo se invertirá tiempo en aquellas subregiones que posiblemente sí contengan una cara.

Conjuntamente, con el uso de este algoritmo se pueden realizar reconocimientos faciales en tiempo real para aplicaciones con un mínimo de 2 fotogramas por segundo.

Este algoritmo aplica un clasificador en cascada, entrenado mediante un método de Machine Learning, con concreto Adaboost. Los descriptores empleados son las características Haar, muy adecuadas para la detección facial. Un dato interesante es que puede combinarse con un sistema de reconocimiento facial basado en video, por ejemplo las salas inteligentes

(1) Recuperado de Carlos Esparza Franco (2017). de <https://www.researchgate.net/publication/319958509>

CAPÍTULO 1

que pueden reconocer personas a iniciar automáticamente acciones apropiadas. Otro ejemplo son los sistemas que detectan la fatiga de un conductor, monitorizando las expresiones de la cara y los movimientos de cabeza.

CAPÍTULO 2

En el capítulo anterior se realizó un recorrido sobre los aspectos elementales sobre el reconocimiento facial, su historia y evolución; también se señalaron algunas técnicas más reconocidas en la detección de rostros, así como el proceso de reconocimiento facial (cuatro etapas). También se indagó en cuanto a algoritmos y sus diferencias, a los efectos de analizar posteriormente, cuál es más conveniente a aplicar en el presente trabajo.

En este capítulo, intentaremos vincular el factor “sonido” con la idea de detallar cuál será el campo de acción del reconocimiento facial: salas, espacios cerrados, entre otros.

El acondicionamiento adecuado de una sala o espacio en el cual se reproducen distintas fuentes sonoras, implica dos problemas: por una parte el control del ruido de las fuentes sonoras no deseadas, existentes tanto en el interior como en el exterior del recinto y que producen inmisiones en el mismo, como la búsqueda de una solución para la adecuada distribución del campo sonoro para cumplir un determinado objetivo.

Control del sonido

La reverberación natural en las salas para música no amplificadas es de gran importancia. Éste, es un fenómeno sonoro producido por la reflexión, que consiste en una ligera permanencia del sonido una vez que la fuente original ha dejado de emitir (Accolti, 2013).

Por ejemplo, el tiempo de reverberación preferido es de aproximadamente 2 s para salas de concierto y del orden de 1 segundo para salas donde el principal uso es la palabra hablada, pasando por valores intermedios como el de aproximadamente 1,5 s para salas de teatro.

Para eventos o espectáculos con amplificación, la reverberación natural es un factor totalmente secundario pues habitualmente se agrega reverberación artificial mediante los equipos electrónicos que componen el equipamiento de sonido. Esto sugiere que la reverberación natural de la sala tendría que ser baja para no obstaculizar la estética sonora del espectáculo. Además existen otros factores, por los cuales la reverberación natural debe ser baja, como la claridad musical y la perturbación producida por

reflexiones sonoras que arriban mucho más tarde que el sonido directo (Accolti, 2013).

Un abordaje sobre en este tema, puede ser considerar las restricciones presupuestarias dentro de cierto margen y obtener diseños comprometidos informando el porcentaje estimado de personas que escucharán adecuadamente y el de aquellas que se verán perturbadas por las reflexiones sonoras y/o que percibirán el espectáculo con una inteligibilidad de la palabra o claridad musical no del todo buena.

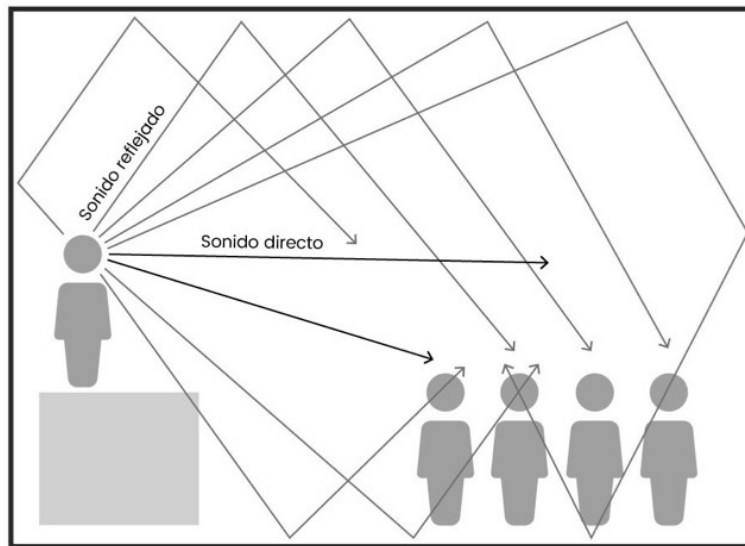


Gráfico de pérdida de inteligibilidad dentro de un recinto

En líneas generales un factor decisivo tiene que ver cuando el volumen del local es grande y las superficies de sus paredes, suelo y techo son lisas y no absorbentes. El ruido rebota en todas las superficies, sobreponiéndose las ondas sonoras entre ellas. Esto amplifica el sonido produciendo ruidos molestos y genera ecos que dificultan la comprensión auditiva.

Es importante saber que para evitar esta situación, se deben implementar elementos absorbentes que corrijan el efecto rebote. Lo más sencillo es aprovechar el techo suspendido, ya que con un buen diseño se puede conseguir romper la onda de sonido y con un buen material se puede absorber gran parte del ruido.

En cuanto a la duración, el tiempo de reverberación depende directamente del volumen e inversamente de la absorción sonora.

Por otra parte, si se tiene en cuenta a las salas amplias que tienen gran volumen de espacio, la extensión de la reverberación tiende a ser alta si no se aplica la absorción necesaria. También cabe considerar la absorción que representa la propia audiencia en caso de sala llena. Particularmente la audiencia no absorbe sonido en frecuencias bajas y es justamente en ese rango donde el estudio acústico debe enfocar un gran esfuerzo. Además los eventos que normalmente se desarrollan en este tipo de salas generan una gran energía sonora en bajas frecuencias que, de no ser absorbida, perjudica la claridad musical (Knauf Blog, 2019).

Acústica interior de la sala

La acústica dentro de la sala se encuentra relacionada con el ruido de fondo: este es otro de los indicadores de la claridad y la inteligibilidad de la música y la palabra respectivamente.

En las salas de gran amplitud no es una de las cuestiones más importantes debido a que en general la música popular genera un nivel sonoro elevado y no posee grandes cambios dinámicos, es decir, ese nivel es más bien constante y generalmente no se evidencian pasajes musicales en planos sonoros de bajo nivel.

En cambio, las salas más pequeñas tendrán problemas más importantes que las de mayores dimensiones a la hora de acomodar contenidos de baja frecuencia.

Generalmente en grandes eventos de música popular el mayor ruido de fondo es generado por los mismos espectadores. Pero es deseable asegurar que no sea demasiado alto el ruido de fondo debido al equipamiento del edificio (por ejemplo el ruido de generadores eléctricos de ventiladores y también el aerodinámico generado en los conductos de inyección y propagados por los difusores).

Un criterio adecuado para el ruido de fondo debido al equipamiento es el sugerido por la norma IRAM 4070/2008 para estudios de TV y salas de música, no solo como sala de música pero también teniendo en cuenta que los espectáculos que se desarrollan en este tipo de salas muchas veces son

trasmitido por TV. Este criterio corresponde a un perfil de ruido entre NC 20 y NC 25 que corresponde aproximadamente a entre 30 dBA y 35 dBA.

La psicoacústica

Entender la psicoacústica es fundamental en este trabajo ya que el diseño de un sistema de reconocimiento facial aplicado al control del sonido en una sala, evento o conferencia requiere tener presente que el público receptor. Este público, tiene características heterogéneas a la hora de oír e interpretar el sonido, de ahí la importancia de adentrarnos en el tema.

De acuerdo a las fuentes consultadas, han señalado que la audición consiste en la capacidad de percibir e interpretar las vibraciones moleculares del medio externo. Relacionado a ello, se encuentra la psicoacústica, es decir una rama de la psicofísica que estudia la relación existente entre las características físicas de un estímulo sonoro, y la respuesta de carácter psicológico que él mismo provoca en un sujeto.

Es una disciplina esencialmente empírica, y sus conclusiones se obtienen a partir del análisis estadístico de los resultados de experimentos que buscan medir la respuesta subjetiva de sujetos a estímulos de propiedades físicas cuantificadas (Wikipedia, 2020).

Tal como lo señalan en Control Rooms Summit (2017) así como la acústica se dirige al examen, uso y aprovechamiento del sonido, la psicoacústica, profundiza el estudio de las limitaciones y ventajas del mismo dentro del sistema auditivo humano: especialmente en cuanto al funcionamiento del oído, umbrales de percepción, enmascaramiento, distorsión, ecolocalización, etc.

Desde el punto de vista de las aplicaciones que tiene la psicoacústica, se encuentran: desarrollo de software y procesamiento de señal mejorando y creando formatos de compresión y novedosas aplicaciones para las ampliamente adoptadas estaciones de trabajo digital de audio, diseño de sistemas sonoros de alta calidad, un sinfín de implementaciones dentro de la música (ya que permite a ingenieros y músicos generar nuevas y definidas experiencias auditivas manejando con precisión y conocimiento los fenómenos de enmascaramiento de frecuencias no deseadas de ciertos instrumentos, como también realizando las armoniosas) y diseño arquitectónico de salas y teatros ya que no sólo es ventajoso tomar en

cuenta los factores físicos del ambiente, sino el hecho de que tanto el oído como el cerebro son parte activa de la experiencia auditiva de toda persona (Control Rooms Summit, 2017).

Estos datos señalados resultan relevantes ya que la psicoacústica podría ser sumamente útil al momento del diseño del sistema de reconocimiento facial, la programación de algoritmos de por medio, para encauzarlo dentro del plano de la reproducción del sonido.

Además, la psicoacústica funciona de manera diferenciada en personas adultas mayores, o personas que padecen de síndrome de autismo, mostrando distintos niveles de sensibilidad auditiva.

ANTECEDENTES

Mugeetion

En la investigación de Stella Koh & Yadegari (2018), titulada “Mugeetion: Musical Interface Using Facial Gesture and Emotion” se expone que las personas sienten emociones cuando escuchan música. Sin embargo, las emociones no son objetos tangibles que pueden ser explotados en el proceso de composición musical ya que son difíciles de capturar y cuantificar en algoritmos. Mugeetion, diseñada para capturar instancias de estados emocionales de los gestos faciales de los usuarios y para retransmitir esos datos a características musicales asociadas. Mugeetion puede traducir datos cualitativos de estados emocionales en datos cuantitativos, que pueden utilizarse en la generación de sonido.

El rostro como dispositivo

Por otro lado, el trabajo de Gutiérrez De Angelis (2017), titulada “El rostro como dispositivo”. De la antropometría a la imagen biométrica” muestra cómo la evolución de la importancia del rostro ha tenido a lo largo de la historia, un papel relevante para distintas disciplinas las cuales se han dedicado a atender este fenómeno, desde una perspectiva cultural, antropométrica y desde determinados dispositivos, los cuales han impuesto articulaciones ópticas espaciales (como lo fue en su momento el caso de la fotografía). Los aportes más relevantes de este trabajo residen en contribuir desde una mirada histórica, con autores reconocidos en materia de técnicas de imagen y rostro, tales como Bertillo, Darwin, y Mitchell, entre otros. Además, entre los temas abordados se encuentran la reconstrucción de rostro de personas que han habitado en tiempos pasados, dispositivos utilizados por los teólogos, pasando por la cultura visual donde diversas tecnologías actuales de la imagen, como la fotografía métrica, los rayos X o los avances de los estudios sobre ADN han permitido develar lo que el cuerpo humano puede decir y ser capaz de mostrar.

MyoSpat

Por otra parte, “MyoSpat: A hand-gesture controlled system for sound and

light projections manipulation” realizado por Balandino, Di Donato, Dooley, James & Hockman, Jason (2017), presentan a MyoSpat, un Sistema interactivo que permite a los artistas controlar las proyecciones de luz y sonido por medio de gestos con las manos. Este programa está diseñado y desarrollado utilizando el brazalete Myo como dispositivo de entrada y Pure Data como motor audiovisual. El sistema se basa en los principios de interacción hombre-computadora; específicamente, computación tangible y diseño de interacción sonora y musical incorporada (MiXD). Este documento cubre una descripción del sistema y su diseño de retroalimentación audiovisual. Finalmente, evaluamos el sistema y su uso potencial en la exploración de principios de interacción corporal, sónico y musicales en diferentes contextos multimedia.

Encuesta de literatura del sistema automático de reconocimiento facial e implementación basada en Eigenface

Otro aporte a los antecedentes tiene que ver con el trabajo llevado a cabo por Shariful Islam Ferdous & Sabbir, Ahmed (2010), el cual se titula “Encuesta de literatura del sistema automático de reconocimiento facial e implementación basada en Eigenface”; la misma, se propone como objetivo es proporcionar una revisión actualizada de la literatura existente, y en segundo lugar, es revelar algún término técnico sobre cómo una computadora puede reconocer una cara (principalmente los algoritmos) y el intención final de dar a conocer la oportunidad de investigación para principiantes y varias perspectivas del campo del reconocimiento facial. Para proporcionar una encuesta completa, se partió de los estudios psicofísicos relevantes sobre cómo el ser humano realiza este trabajo. Luego, pase a las técnicas de reconocimientos existentes y presente descripciones detalladas de todas las técnicas. Además, puse algunos estudios comparativos entre sistemas se explicó alguna dificultad de preprocesamiento (iluminación, variación de pose, calidad de imagen, etc.) y su definición para el sistema de reconocimiento facial. Finalmente intentamos implementar un sistema AFR basado en el método Eigenface donde se ha agregado un filtrado de ruido.

Sonify Your Face

Conforme a la búsqueda de antecedentes relacionados a la temática escogida se encuentra un trabajo titulado “Sonify Your Face: Facial

Expressions for Sound Generation” por Velenti, Jaimes & Sebe (2010), presenta una novedosa herramienta de creatividad visual que reconoce automáticamente las expresiones faciales y rastrea los movimientos musculares faciales en tiempo real para producir sonidos. El módulo de reconocimiento de expresiones faciales detecta y rastrea una cara y genera un vector de características de movimientos de ubicaciones específicas en la cara. El vector de características se usa como entrada a una red bayesiana que clasifica las expresiones faciales en varias categorías (por ejemplo, enojado, disgustado, feliz, etc.). Los resultados de la clasificación se utilizan junto con el vector de características para generar una combinación de sonidos que cambian en tiempo real según las expresiones faciales de la persona.

Facing the Music a Facial Action Controlled Music

En otra indagación sobre antecedentes se encontró “Facing the Music a Facial Action Controlled Music” por Lyons, Michael & Tesutani, Nobuji (s/f), el cual se relaciona con el presente estudio ya que, describe un novedoso controlador musical que adquiere en vivo entrada de video de la cara del usuario, extrae las característica facial parámetros utilizando un algoritmo de visión por computadora y convierte estos a efectos musicales expresivos. En suma, el controlador permite que el usuario modifique música sintetizada o filtrada por audio en tiempo real en el movimiento del rostro. El trabajo mencionado, ha logrado demostrar que el sistema de usos por computadora por medio de la visión ha logrado convertir la acción facial en efectos musicales.

En definitiva, se pudo observar que la literatura en torno a nuestro objeto de estudio es capaz de ofrecer diversos ángulos el cual podrán enriquecer, los resultados llegado el momento.

CAPÍTULO 3

Proyección del dispositivo

En apartados anteriores se trabajó sobre los aspectos conceptuales necesarios y elementales, en vista a considerar el diseño y elaboración de mi dispositivo, el cual deberá tener la capacidad de controlar una o varias fuentes de sonido y de detectar al público de un recinto determinado haciendo uso del reconocimiento facial, haré un desarrollo de los componentes que ya tengo materializados para dicho dispositivo, así como aquellos que aún están en etapa de diseño.

Como primera medida, debo aclarar que la versión más reducida del dispositivo que me encuentro desarrollando constará de un bafle de sonido con una cámara de video insertada en su interior, y ambos elementos estarán conectados a una computadora (o procesador dedicado), la que se encargará de procesar y gestionar las señales de audio a través de una interfaz externa, con múltiples entradas y salidas. La conexión entre el bafle y la computadora será por medio de un cable de señal de audio, en tanto que la cámara se conectará inalámbricamente, para evitar de esta forma los cableados innecesarios.

Las versiones más complejas del dispositivo constarán de mayor número de bafles y cámaras, estas cantidades estarán determinadas por el tipo de escucha que se encuentre instalado en la sala que se quiera controlar.

Cabe aclarar que los diferentes tipos de escucha más comunes que podemos encontrar en los recintos en los cuales se realizan diferentes tipos de obras sonoras son los siguientes: un solo bafle para una escucha del tipo monoaural, dos bafles para un sistema estéreo, o más bafles para situaciones más complejas, como bien podría ser el caso de un refuerzo sonoro para varias áreas simultáneas (palcos, Side Fills, Out Fills, u otras.) de un recinto de grandes dimensiones. También podrían usarse mayor cantidad de bafles para simulaciones del tipo 5.1 o 7.1.

Los bafles que están siendo fabricados también permiten sumarse en un sistema más complejo del tipo Line Array, tal como se ve en esta imagen.

Sistema Line Array



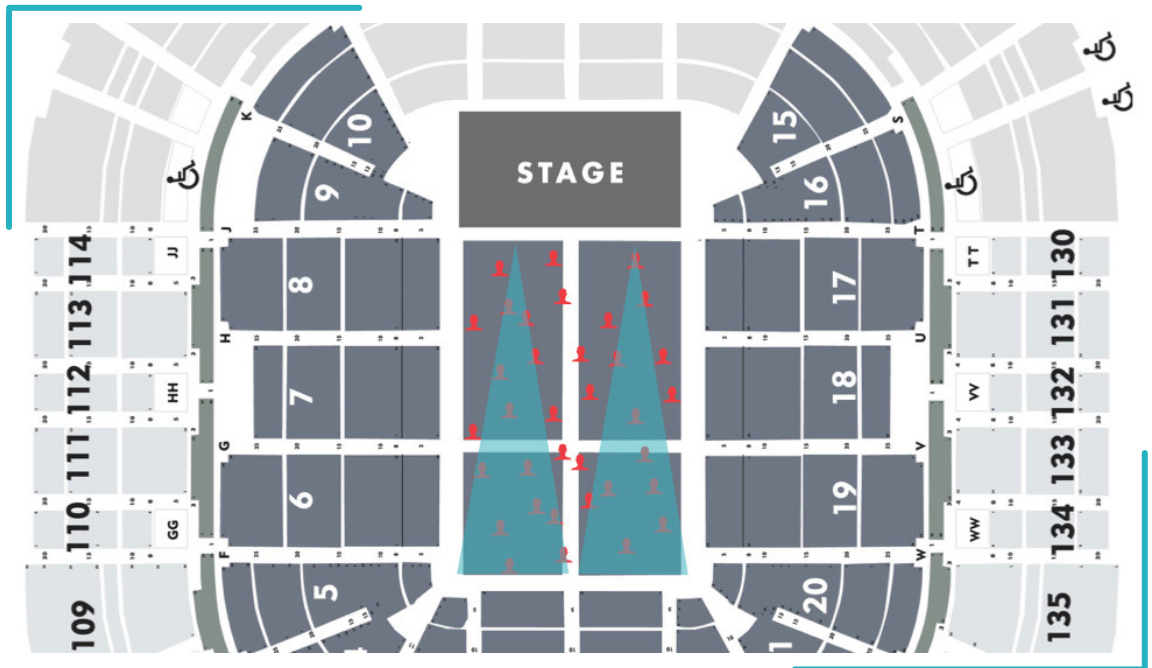
Sistema del tipo Line Array (arreglo en línea) - 3 Mid-Hi por lado + Sub

En futuras pruebas se evaluará si la lectura de los rostros y la programación tendrán mayor efectividad ubicando la cámara dentro del propio dispositivo de audio (bafle), tal como se comentó más arriba, o ubicando la misma en puntos ajenos a dicho bafle.

Consideraciones en cuanto a la característica de las cámaras de video

Cada bafle o conjunto de bafles deberá estar vinculado con una cámara de video, con una resolución mínima de 1920 x 1080 mega pixels (estándar FULL HD).

Estas cámaras deberán tener un ángulo de visión no menor a los 90° de ángulo, tanto vertical como horizontal. De este modo, se podrá captar la totalidad del espacio ocupado por el público, teniendo en cuenta que al tratarse en su mayoría de sistemas de sonido estereofónicos (entiéndase: sistemas de dos fuentes, una encargada de reproducir el canal izquierdo de audio y la otra encargada del canal derecho), y que mínimamente habrá una cámara de video en cada una de dichas fuentes (bafles), un sistema básico contará con al menos 2 cámaras. Ambas cámaras deberán captar al público en su totalidad, dividiéndose el espacio como lo muestra la siguiente imagen.



Ubicación y ángulo mínimo de las cámaras de video dentro de un recinto, y para un sistema estereofónico.

En lugar de utilizar varias cámaras para tomar más de un sector de una sala se puede optar por una sola cámara con movimiento, es decir, una que vaya creando un “barrido” de esos sectores, tantas veces por minuto como sea necesario. De esta manera se podrá minimizar no sólo el número de cámaras, sino también la cantidad de fuentes de video que ingresan al sistema de procesamiento, resultando beneficioso también en lo referido al presupuesto, ya que el costo de cada cámara es considerable.

En este último caso bastará entonces con “mapear” previamente la/s zona/s en cuestión, es decir, antes de que ingrese el público, para delimitar los sectores que pueden presentar variaciones en cuanto a número de personas o público, para luego poder cancelar la emisión de sonido de ese sector en particular, o controlarlo según sea el caso.

También se espera que cada baffle pueda cambiar su patrón o lóbulo de radiación, para así poder controlar cuán abierto o cerrado es el ángulo de cobertura, tanto horizontal como vertical. De esta forma se podrá direccionar la emisión de sonido solo a sectores con público, y anular la que se dirige a zonas sin audiencia. Esto mejorará notablemente la inteligibilidad en el lugar, principio acústico mencionado en los apartados anteriores.

Cabe aclarar que para poder modificar el patrón de radiación de cada fuente de sonido independientemente, y así la del resto de los sistemas, es necesario que cada una cuente con al menos 2 altavoces (parlantes), porque

así y solo así se podrá hacer un corrimiento de fase (diferencia de llegada temporal) entre ambos parlantes, lo que dará como resultado diferentes patrones de directividad.

Construcción del dispositivo - etapas ya materializadas

Fabricación de los bafles (fuentes de sonido)

Los bafles ya se encuentran en etapa de fabricación, y fueron diseñados por quien suscribe, con algunas colaboraciones puntuales. Los mismos constan de 2 parlantes de 10" y un driver (transductor para frecuencias agudas) de 1,5", y permite ser alimentado como sistema de 2 o 3 vías. En la siguiente imagen se puede ver su aspecto terminado.



Sistema de altavoces - Modelo p210 (pasivo adaptable a 2 o 3 vías)

Este primer modelo fabricado desarrolla una potencia de 700 watts RMS por unidad, y cubre un rango de frecuencias de entre 30 y 20.000 hertz. Una sola unidad puede sonorizar una sala con aproximadamente 100 personas.

Las características generales del gabinete son las siguientes:

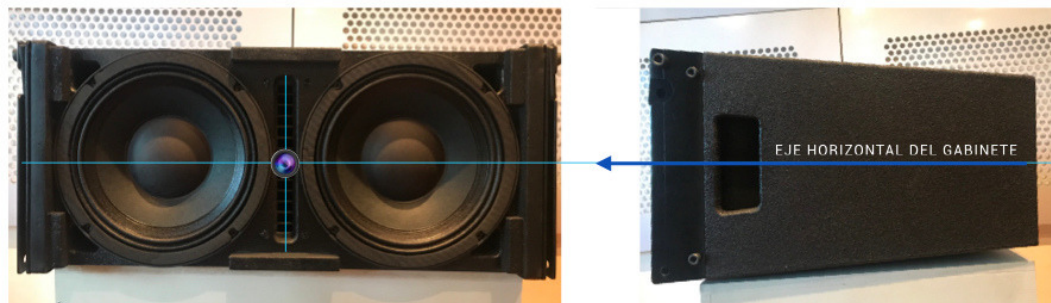
- Sistema pasivo de 3 Vías
- Componentes fabricados íntegramente en Italia (no tiene componentes Chinos)
- Gabinetes Mid-Hi de 700 Watts RMS (REALES)
- Sub Low Pasivo de 1000 Watts RMS (REALES)
- Conectores de calidad profesional
- Pintura de alto impacto

Especificaciones técnicas:

- Gabinete Medios-Agudos
- Parlantes X 2: Faital Pro (10" / 300W RMS / 97dB / 8 ohm) + Driver X 1: Faital Pro (Bobina 1,50" / 75W / 109 dB / 8 ohm)
- Sub Low
- Parlante X 1: Faital Pro (18" / 1000W RMS / 99 dB / 8ohm)
- Respuesta de frecuencia del sistema: 30 Hz / 20.000 Hz.
- SPL Frecuencias baja Continua: 99 dB
- SPL Frecuencias Altas Continua: 109 dB
- Cobertura Horizontal (-6 dB): 120 °
- Cobertura Vertical (-6dB): Dependiendo del Array Crossover.
- Frecuencias de cruce ideal: 98 Hz

En el centro del gabinete se ubicará una cámara de video con las características mencionadas anteriormente, quedando en una ubicación "ON AXIS", es decir, en eje con el centro horizontal de la caja, tal como lo muestra la siguiente imagen. Esta cámara se podrá tomar como única fuente de video o podría sumarse con otras, dependiendo de la situación.

La ubicación ON AXIS tiene como propósito hacer coincidir la dirección de la emisión sonora del gabinete de audio con el centro de captación de la cámara, y que exista correlación espacial entre ambos objetos.



Ubicación On Axis de la cámara dentro del bafle

El procesamiento se hará con una computadora portátil, conectada a una interfaz de audio de múltiples entradas y salidas.

Esta computadora albergará los softwares dedicados para el reconocimiento facial (Python) y para el procesamiento del audio (Max), así como las múltiples aplicaciones adicionales necesarias para el correcto funcionamiento de todo el sistema.

Tanto la computadora como la interfaz de audio externa deben ser compatibles en cuanto a respuesta en frecuencia de trabajo (20-20.000hz) y profundidad de bits (24 bits como mínimo).

Importante: la latencia (delay) de ambos componentes no debe superar el valor de 6 milisegundos, ya que de otra manera se verían comprometidos los tiempos de arribo a cada dispositivo emisor.

Construcción del dispositivo - etapas a completar

– Programación de captura de datos desde las cámaras en entorno dedicado (Python)

Será necesario el diseño, programación e instalación de un software que sea capaz de captar a un grupo numeroso de personas dentro de la sala. También será indispensable realizar pruebas de reconocimiento biométrico

de personas con problemáticas en la audición, como por ejemplo personas de avanzada edad, pacientes con trastorno del espectro autista, quienes hayan sufrido algún trauma acústico, etc.

La programación se realizará en la plataforma Python, y será realizada por una persona entendida en esta temática en particular, la cual escapa a la formación que poseo en el campo del audio.

Además de Python debemos hacer uso de las Apps “OpenCV” y “Face Recognition”, la primera permite solo detectar donde hay un rostro dentro de una imagen, y la segunda permite reconocer rostros desde una base de datos previamente cargada (fotografías de referencia). OpenCV utiliza el algoritmo de Viola & Jones para detectar rostros, que desde su publicación en 2001 es probablemente el algoritmo por excelencia para hacer este tipo de tareas. Este algoritmo está pensado para detectar objetos concretos, ya sean rostros, coches, gatos, plátanos...

Para poder detectar objetos se necesita un clasificador. Podemos imaginar los clasificadores como una especie de “plantilla” para detectar un determinado tipo de objeto. Este clasificador es entrenado proporcionándole cientos de imágenes de ejemplo del objeto que se quiere detectar.

Cuando se quieren detectar objetos, el algoritmo aplica una serie de clasificadores (una “cascada de clasificadores”) sobre una región de interés en la que considera que podría hallarse el objeto en cuestión. Si todos los clasificadores acaban dando positivo, el algoritmo considera que hay una coincidencia y se ha encontrado el objeto deseado. En nuestro caso, un rostro.

Reconocimiento facial con OpenCV y FaceRecognition

Una vez encontradas las caras de una imagen resta identificar a las personas que aparecen y relacionarlas con sus nombres, los que se cargaron en la base de datos.

OpenCV es una muy buena librería de visión por computadora, y proporciona herramientas muy efectivas para detectar rostros. Sin embargo, no dispone de funciones para identificar fácilmente a personas. Por lo tanto, será necesario utilizar la librería Face_Recognition que está específicamente diseñada para hacer reconocimiento facial de humanos.

Un desafío en cuanto a programación es lograr una buena captación de los rostros no solo en situaciones donde la cámara tome a un individuo o poca cantidad de personas cercanas a ésta, sino que también debe conseguirse una eficacia en cuanto a conteo e identificación aún en públicos multitudinarios, en donde se debe tener en cuenta también el movimiento de las personas en su propio eje, así como los desplazamientos relativos dentro del recinto. Se entiende que este objetivo es uno de los más desafiantes del proyecto.

Podemos aclarar que el reconocimiento de rostros con una o pocas personas fue realizado con una alta efectividad y un muy bajo margen de error, ya que a nivel de requisitos técnicos así como de código de programación son alcanzables. También sabemos que el reconocimiento de personas que se encuentran a cientos de metros de la cámara, en espacios muy amplios, o en un entorno de muchedumbre es posible solo con tecnología de última generación y códigos de programación súper exclusivos, propios de capitales privados dedicados por ejemplo a la bioseguridad, y con presupuestos que rozan cifras millonarias.

Será nuestro (el de la persona que me asista con la programación y el mío) principal desafío lograr resultados similares pero con un presupuesto y equipamiento acorde a nuestra realidad actual. Por las pruebas que comenzamos a hacer no descartamos que esto sea posible.

The screenshot shows a Python IDE with a file explorer on the left and a code editor on the right. The file explorer shows a Django project structure with folders like bin, conf, contrib, core, db, dispatch, forms, http, middleware, shortcuts, template, templatetags, test, utils, and views. The code editor shows the content of loader.py, which includes methods for __init__, reload, make_origin, and find_template_loader.

```

74 def __init__(self, display_name, loader, name, dirs):
75     super(LoaderOrigin, self).__init__(display_name)
76     self.loader, self.loadname, self.dirs = loader, name, dirs
77
78 def reload(self):
79     return self.loader(self.loadname, self.dirs)[0]
80
81 def make_origin(display_name, loader, name, dirs):
82     if settings.TEMPLATE_DEBUG and display_name:
83         return LoaderOrigin(display_name, loader, name, dirs)
84     else:
85         return None
86
87 def find_template_loader(loader):
88     if isinstance(loader, (tuple, list)):
89         loader, args = loader[0], loader[1:]
90     else:
91         args = []
92     if isinstance(loader, basestring):
93         module, attr = loader.rsplit('.', 1)
94         try:
95             mod = import_module(module)
96         except ImportError as e:
97             raise ImproperlyConfigured('Error importing template source loader: %s' % module)
98         try:
99             TemplateLoader = getattr(mod, attr)
100         except AttributeError as e:

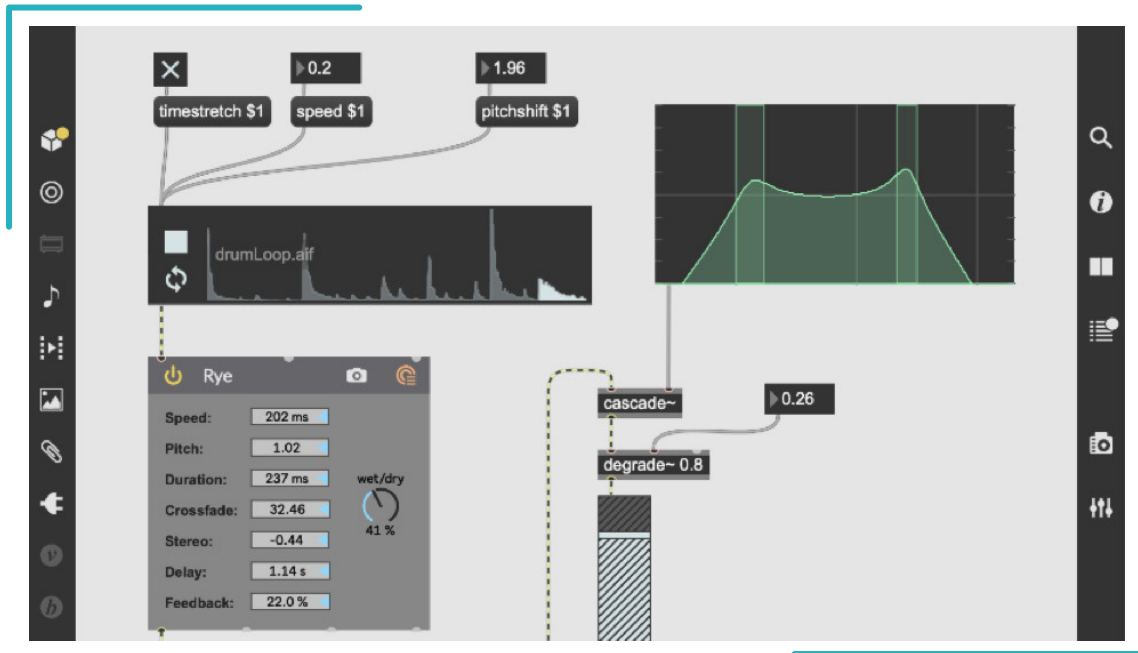
```

Entorno de programación Python. Fuente: recuperado de w3big.com.

Creación de patch de audio en MAX

Será necesaria la creación de un patch (diseño personalizado de conexasión interno que vincula las entradas, el procesamiento y las salidas de audio dentro de ese entorno) dedicado, que permita recibir las señales de audio del sistema en cuestión, en su mayoría son señales L y R del estéreo, y que dentro de este software sea posible su procesamiento, haciendo uso de los datos obtenidos desde las cámaras de video, para que por último sean manipuladas las salidas de audio a los diferentes canales (L y R, 5.1, 7.1, etc.)

Importante: Max es uno de los entornos de desarrollo gráfico con código abierto para audio y multimedia para más potentes del mercado, siendo muchas veces el paso previo al “enlatado” del programa en un microcontrolador autónomo, como puede ser el Atmega 328p.



Fuente: recuperado de cycling74.com, entorno gráfico estándar nativo de Max

Una alternativa mucho más visual al entorno de la interfaz estándar de Max (imagen anterior) es la Api “Max for Live”, que permite la vinculación de este programa con el ya conocido software de producción Ableton Live. Los comandos y parámetros están mucho más claros, dado que este entorno se parece más a los secuenciadores por software más usados del mercado. Esta interfaz se puede ver en la siguiente imagen.



Fuente: recuperado de cycling74.com, entorno gráfico propio de Max for Live

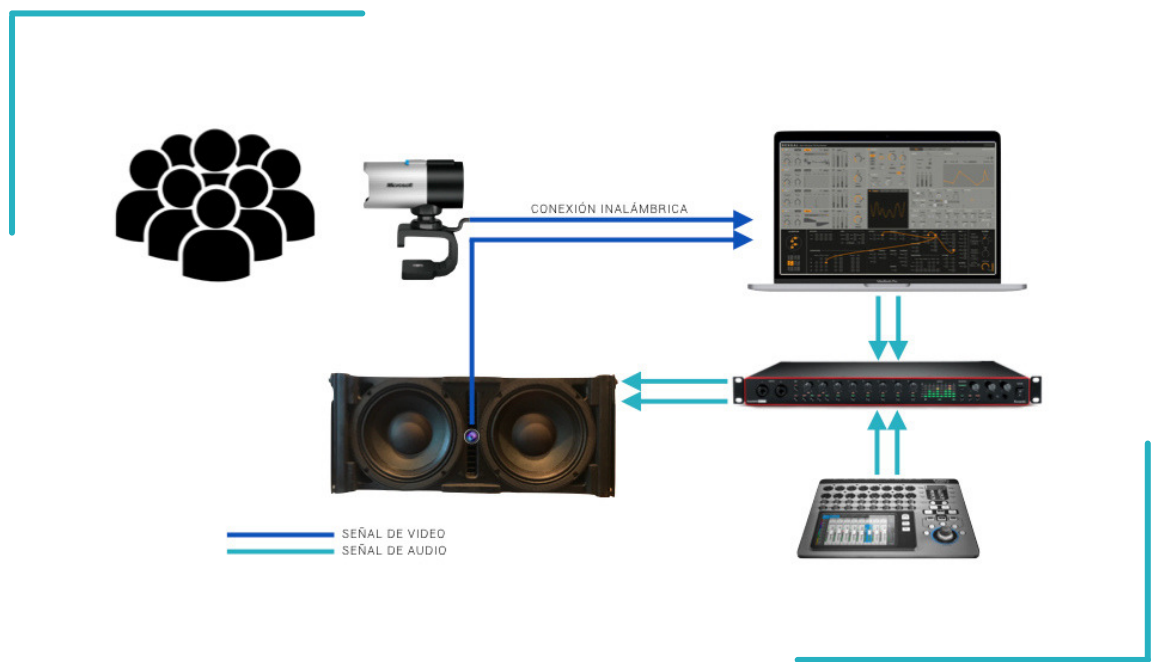
Consideraciones para la instancia práctica

En la etapa de prueba de campo se comenzará a testear el comportamiento del dispositivo en su versión reducida, y serán necesarios una sala y un público acordes al tamaño del bafle que se está probando. El público servirá de muestra para la detección y reconocimiento de los rostros, para así llevar a cabo los ajustes de la programación de los softwares.

Quedará también la medición y comparación del comportamiento de las cámaras de video dentro de los bafles o en otras ubicaciones del recinto, pudiendo así comprobar en cual de estas dos situaciones, o en la conjunción de ambas, es más fácil corregir y promediar el ángulo de radiación (patrón de radiación) de un bafle o sistema compuesto, como es el caso de los ya mencionados Line Array.

Considerando todos estos puntos se estará en mejores condiciones para llevar adelante el diseño definitivo del dispositivo, aunque como se puede observar, cada etapa modifica notablemente a las otras.

El sistema terminado estará interconectado como lo muestra la siguiente imagen.



Dirección de las correspondientes señales a sus respectivos destinos

CONCLUSIÓN

Al inicio del trabajo se consideró analizar y explorar las necesidades para la elaboración de un dispositivo capaz de brindar control y procesamiento de sonido utilizando los datos de algoritmos usados para el reconocimiento facial.

Los fundamentos que me llevan a realizar este proyecto no solamente tienen relación directa con mi profesión, sino con la posibilidad de ofrecer un servicio que sea capaz de dar solución a quienes enfrentan dificultades auditivas, o personas que tienen problemas en situaciones de eventos, obras sonoras, salas de conferencias, etc.

Por lo tanto, realizar un aporte para adaptar y moderar la emisión de sonidos, mejorar la inteligibilidad del sonido, identificar parámetros gestuales de grupos constituye objetivos necesarios a intervenir para mejorar la experiencia del público en general.

En tal sentido, a lo largo de este trabajo se hizo un recorrido necesario sobre el campo enorme del reconocimiento facial, sin intentar agotar la indagación en el tema.

No caben dudas que vincular el reconocimiento facial con el control del sonido aplicado no sólo de modo individual, sino grupalmente a un colectivo de personas dentro de una sala o evento, representa un gran desafío en la realidad.

Es por ello que se tuvo presente el análisis de las técnicas de reconocimiento, los algoritmos más conocidos, así como también el estudio de la acústica interior, y la psicoacústica. Llevar adelante el proyecto requiere la interiorización sobre estos temas, ya que constituyen la plataforma inicial para posteriormente materializar la idea inicial del trabajo.

BIBLIOGRAFÍA

Reconocimiento facial

- Aldana, R. (2017). 6 Emociones básicas, latidos de un lenguaje universal. La mente es maravillosa. Obtenido de <https://lamenteesmaravillosa.com/6-emociones-basicas-latidos-de-un-lenguaje-universal>
- Espinoza, David & Jorquera, Ignacio (2015). Reconocimiento Facial. Escuela de Ingeniería Informática.
- Calli Olvea, J. (2015). Reconocimiento facial basado en el algoritmo eigenface. Lima: Universidad "Andina Néstor Cáceres Velásquez".
- Chávez Rodríguez, J. (2017). Sistema de Control de Acceso empleando Reconocimiento Facial en entornos controlados. Santiago de Cuba: Universidad de Oriente. Hernández, G. (2010). Estudio de técnicas de reconocimiento facial. Barcelona: Universidad Politécnica de Cataluña. Recuperado el 02 de 2020, de https://upcommons.upc.edu/bitstream/handle/2099.1/9782/PFC_Roger_Gimeno.pdf
- Balandino, Di Donato, Dooley, James & Hockman, Jason (2017). MyoSpat: A hand-gesture controlled system for sound and light projections manipulation. Creative Commons Attribution License 3.0 Unported. Recuperado el 07-10 Del 2020 de https://www.researchgate.net/publication/320427332_MyoSpat_A_hand-gesture_controlled_system_for_sound_and_light_projections_manipulation
- G. De Angelis, Marina. "El rostro como dispositivo. De la antropometría a la imagen biométrica", e-imagen Revista 2.0, N° 4, Sans Soleil Ediciones, España-Argentina, 2017, ISSN 2362-4981
- Kochetkove, K. (2016). Hombre contra la máquina: reconocimiento facial. Recuperado el 2020, de <https://www.kaspersky.es/blog/how-facial-recognition-works/8564/>

- López, Miguel Ángel (2014). Sistema de Reconocimiento Facial Mediante Técnicas de Visión Tridimensional. Centro de Investigaciones en Óptica. Lyons, Michael & Tetsutani, Nobuji (s/f), Facing the music: a facial action controlled musical interface. ART Media Integration and Communication Research Laboratories. Kyoto, Japan.
- Stella Koh, E., & Yadegari, S. (2018). Mugeetion: Musical Interface Using Facial Gesture and Emotion. San Diego, USA. Recuperado el 05 de 02 de 2020, de <https://arxiv.org/pdf/1809.05502.pdf>
- Valenti, R., Jaimes, A., & Sebe, N. (2010). Sonify Your Face: Facial Expressions for Sound Generation. The Netherlands: University of Amsterdam. Recuperado el 05 de 02 de 2020, de <https://www.semanticscholar.org/paper/Sonify-your-face-facial-expressions-for-so-und-Valenti-Jaimes/2952482c50ffac6f9104109a50c5ea53f779d0e5>
- Nerea, Navio (2017). Sistema de identificación de personas mediante reconocimiento facial aplicado a videovigilancia. Gandia.
- Byung - Chul Han, Shanzai. (2016/2017). El arte de la falsificación y la deconstrucción en China, Caja Negra, Bs. As.
- Rearte, Manuela, Sánchez Martín (2011). Introducción a la Inteligencia Artificial. Universidad Nacional de Rosario, Argentina.
- Rehm Matthias, Bee Nikolaus, André Elisabeth (2007). Wave Like an Egyptian - Accelerometer Based Gesture Recognition for Culture Specific Interactions.
- Shariful Islam Ferdous, Sabbir Ahmed (2010). Literature survey of automatic face recognition system and eigenface based implementation.

Metodología

- Stagnaro, Daniela (2010). El proyecto de Intervención. Pre-textos, Barcelona.

Control del sonido

- Accolti, Ernesto (2013). Salas para grandes eventos de música amplificada. Proyecto Acústico. Rosario, Argentina.
- (Knauf Blog, 2019). Cómo evitar la reverberación y ecos. Recuperado de <http://blog.knauf.es/reverberaciones-y-ecos-como-evitarlos/#:~:text=Para%20evitar%20esta%20situaci%C3%B3n%20se,absorber%20gran%20parte%20del%20ruido.>

BIO

Me desempeño como técnico de sonido para shows en vivo, realizando también las tareas de venta e instalación de sistemas de Audio Profesional, tanto para estudios de grabación como para instalaciones fijas.

En el último año me encuentro diseñando y fabricando gabinetes acústicos de diferentes modelos, cómo ser convencionales (full range) y Line Arrays, incursionando también en diseños Hi Fi y Home Cinema.

Previo a cursar la carrera de Artes Electrónicas completé estudios en el ámbito de la producción musical e ingeniería de grabación, así como en las áreas de arreglos y orquestación musical.

La elección del tema presentado para mi proyecto tiene relación directa con mi profesión, y radica en el especial interés por el potencial que habría en la fusión de dispositivos y/o software dedicados para la creación y reproducción sonora y las múltiples variables de control que pudieran entregar las herramientas y algoritmos propios del reconocimiento facial. Permitiendo con dicha fusión la generación de nuevas técnicas de producción, en tanto obra instalaciones, dispositivos de uso comercial o variado material sonoro, a la vez que innumerables formas de control y procesamiento de señales de audio mediante los rasgos faciales de un espectador en particular o de todos aquellos que formen parte de un público o audiencia determinada.

Tengo muchísimo interés en diseñar y fabricar sistemas de audio capaces de interactuar con los espectadores a partir del uso de las técnicas mencionadas.

