

UNIVERSIDAD SANTO TOMÁS
FACULTAD DE INGENIERÍA ELECTRÓNICA
GRUPO DE ESTUDIO Y DESARROLLO EN ROBÓTICA GED



“Sistema de reproducción de mensajes multimodales para la cabeza de un robot social”

Propuesta Convocatoria FODEIN 2018

Ing. Carlos Saith Rodríguez Rojas MSc
Ing. Eyberth Rolando Rojas Martínez

11 de septiembre de 2017

Título

Sistema de reproducción de mensajes multimodales para la cabeza de un robot social

Duración

9 meses

Lugar de ejecución

Universidad Santo Tomás, Bogotá D.C, Sede Principal.

Investigador principal

Ing. Carlos Saith Rodríguez Rojas MSc.

Co-investigador(es)

Ing. Eyberth Rolando Rojas Martínez.

Datos generales					
Programa	Facultad	Línea activa	Línea medular	Campos de acción institucional	Grupo de investigación
Ingeniería Electrónica	Ingeniería Electrónica	Inteligencia Computacional Robótica, Visión Artificial	Alberto Magno	Derechos humanos, ciudadanía y construcción de política pública en y para escenarios de paz () Desarrollo tecnológico con apuesta social (X) Desarrollo ambiental sostenible() Cambio educativo y social desde la multi e interculturalidad ()	Grupo de Estudio y Desarrollo en Robótica GED



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@santotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



Institución de Educación Superior sujeta a la inspección y vigilancia del Ministerio de Educación Nacional - SNIES 1704

Equipo de investigación requerido

- **Investigador Principal:** Carlos Saith Rodríguez Rojas, Ingeniero Electrónico de la Universidad Santo Tomás y Magister en Ingeniería de Sistemas y Computadores de la Universidad de los Andes. Docente de tiempo completo de la Facultad de Ingeniería Electrónica desde 2010 y miembro desde 2007 del Grupo de Estudio y Desarrollo en Robótica. Cuenta con numerosos productos de generación de conocimiento de alto impacto, de desarrollo tecnológico y de formación de recursos humanos fruto de mas de 10 proyectos de investigación en calidad de auxiliar, co-investigador e investigador principal.
- **Co-investigador:** Eyberth Rolando Rojas Martínez, Ingeniero Electrónico de la Universidad Santo Tomás y candidato a magister en Ingeniería Electrónica de la Universidad Nacional de Colombia. Docente de tiempo completo de la Facultad de Ingeniería Electrónica desde 2010 y miembro desde 2008 del Grupo de Estudio y Desarrollo en Robótica. Ha participado en más de 8 proyectos de investigación con producción certificada.
- **Asistente de investigación:** Juan Diego Giraldo Rodríguez, graduado en 2017 como Ingeniero Electrónico de la Universidad Santo Tomás con trabajo de grado meritorio, miembro del Semillero de Estudio en Robótica SERO desde 2011 e inscrito para cursar la Maestría en Ingeniería Electrónica con énfasis en Robótica de la Universidad a partir del primer semestre de 2018.
- **Semillero de Investigación:** Semillero de Estudio en Robótica SERO activo desde 2012 y adscrito al Grupo de Estudio y Desarrollo en Robótica. Cuenta actualmente con alrededor de 20 estudiantes de pregrado y se proyectan 4 estudiantes de maestría para 2018.

Alianza estratégica

Gracias a que el Grupo de Estudio y Desarrollo en Robótica GED pertenece a la red mundial *RoboCup* [1], el proyecto cuenta con asesorías académicas y científicas de diferentes universidades a nivel mundial. Algunas con las cuales se ha tenido el mayor contacto en cooperación científica son:

- Carnegie Mellon University, USA, Coral Research Group, Prof. Manuela Veloso PhD.



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



- University of Massachusetts Amherst, USA, Autonomous Mobile Robotics Laboratory, Prof. Joydeep Biswas PhD.
- Universidad Autónoma de México UNAM, México, BioRobotics Laboratory, Marco Negrete PhD.

Resumen de la propuesta

La robótica social trabaja en mejorar la calidad de vida de las personas, ya sea en sus trabajos, hogares o en situaciones de emergencia. Actualmente uno de los proyectos más importantes del Grupo de Estudio y Desarrollo en Robótica de la Universidad Santo Tomás es **MARTHA** (Mobile Autonomous RoboTics Home Assistant) que trata el diseño y desarrollo de un prototipo de robot social acorde a la visión de robótica humanista.

Enmarcada en el proyecto **MARTHA** la propuesta plantea el diseño la implementación de un cabeza para el robot social que le permita la reproducción de manera multimodal de mensajes escritos. La multimodalidad consta de dos aspectos: la emisión de voz correspondiente al mensaje y la gesticulación sobre una cara virtual articulada. Se espera que la integración de un sistema de inteligencia artificial con los recursos de reproducción multimodal permitirán a un robot social un buen grado de naturalidad en el contexto de interacción Humano-Robot. Se proyectan productos de generación de nuevo conocimiento, de desarrollo tecnológico y de formación de recurso humano con la inclusión de un estudiante de maestría como asistente de investigación

Palabras Clave

Robótica Social, Interacción Humano-Robot Natural, Inteligencia Artificial, Aprendizaje de Máquina.

Planteamiento del problema y pregunta de investigación

La robótica social o de servicio es actualmente un campo de investigación en donde se reúnen varias disciplinas desde el diseño mecánico, la inteligencia artificial, la visión artificial hasta la psicología y otras ciencias cognitivas para construir máquinas que interactúen con humanos de forma armoniosa y natural, con el fin de cooperar para solucionar o mitigar problemas de la sociedad. Dichos problemas abarcan entornos in-



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



dustriales, domésticos o médicos.

Uno de los retos mas importantes dentro el desarrollo de robots de servicio es la interacción Humano-Robot natural (*HRI*), la cuál busca que personas y robots interactúen de manera similar a la comunicación entre seres humanos. Según [2] es necesario crear un modelo basado en el procesamiento de información consciente e inconsciente del ser humano; Las personas realizan acciones como la comprensión del lenguaje y la generación de enunciados y gestos de forma consciente sin embargo la naturalidad de las interacciones radica en las acciones inconscientes como el uso de las propiedades del cuerpo, la mirada o los gestos mientras se tiene una conversación.

Específicamente esta propuesta se centra en el problema de *Visual Speech*, el cual consiste en la reproducción de un discurso escrito sobre los recursos de la cara de un robot social con la capacidad de emitir sonidos y realizar gestos faciales. El problema incluye la construcción de un modelo computacional que permita la producción y sincronización de los diferentes fonemas del discurso con los movimientos faciales de manera que se perciba algún grado de naturalidad.

Pregunta de investigación

¿Cómo crear un modelo computacional que permita a un robot social la reproducción multimodal de un mensaje escrito dentro de un contexto de interacción Humano-Robot natural?

Justificación

Los robots día a día se posicionan dentro de nuestras vidas ya sea como asistentes en nuestro hogar y trabajo o como asistente de salud. Cualquiera que sea el contexto el fin ultimo de esta tecnología es el mejoramiento de la calidad de vida de las personas. Específicamente la interacción Humano-Robot es sumamente importante para que la cooperación y/o comunicación entre personas y máquinas sea clara, eficiente, amable y eficaz, consiguiendo que el humano tenga cierta confianza en el robot y pueda delegar tareas que sea incapaz de hacer o que sean peligrosas para su salud. Todas estas tecnologías tendrán un gran impacto social sobre las futuras generaciones ya que estas podrán ocuparse de forma personalizada y en todo momento de la asistencia y tratamiento de personas con discapacidad, cooperación en procedimientos de rescate de victimas, manipulación de sustancias peligrosas, tratamiento de enfermedades físicas y mentales, entre otras.



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



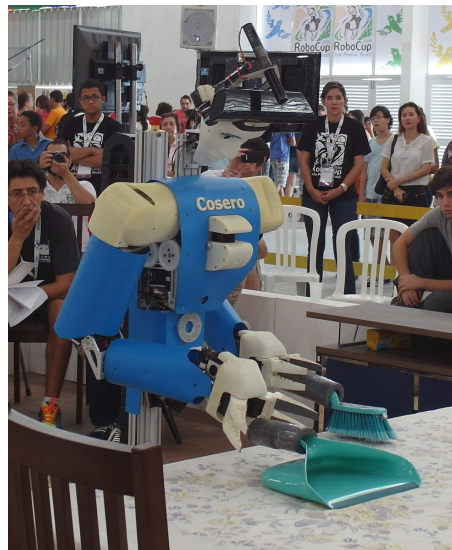


Figura 1: Robot *Cosero*, participante de la Liga RoboCup@Home [6]

En términos investigativos la propuesta se soporta en *A Roadmap for U.S. Robotics: From Internet to Robotics*[3] and *Robotics 2020 Multi-annual Roadmap for Robotics in Europe*[4] cuyos lineamientos o retos en robótica apuntan a la integración de los robots en la cotidianidad de la sociedad, la interacción natural y la cooperación humano-robot.

La propuesta se presenta como una solución al problema de comunicación visual y auditiva del proyecto macro titulado MARTHA (*Mobile Autonomous RoboTics Home Assistant*) concebido por el Grupo de Estudio y Desarrollo en Robótica G.E.D de la Universidad Santo Tomás en el año 2015 con el fin último de crear un robot de asistencia doméstica integral. Todo esto con el soporte, metodología y cooperación académica de la red mundial de investigación *RoboCup*[5], precisamente de la liga *Robocup@Home* cuya competencia mostrada en la Figura 1 es la más importante a nivel mundial de robots de servicio y su objetivo es el desarrollo de tecnología robótica para servicio y asistencia con alta relevancia para futuras aplicaciones domesticas personales.

Objetivo general

Dotar un robot de servicio con recursos de hardware y software que le permitan reproducir en forma de sonido y gestos faciales un mensaje escrito de manera natural.



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



Objetivos específicos

- Construir un prototipo de cara robótica mediante la proyección de un modelo virtual articulado (*Rigging 3D Model*) sobre una máscara semitransparente y la emisión de sonidos por medio de una bocina convencional.
- Implementar un sistema de análisis de texto que permita la extracción y reproducción de una lista de fonemas correspondientes al mensaje escrito.
- Crear un modelo computacional que permita el control de los gestos del robot en coherencia con los fonemas que éste emite favoreciendo la naturalidad de la interacción Humano-Robot.
- Evaluar la eficiencia y naturalidad del sistema en un contexto de interacción Humano-Robot.

Marco Teórico

Interacción Humano-Robot

Basándose en [7], el objetivo de esta sección es dar una introducción al área de interacción humano robot, su objetivo y algunos retos de investigación.

Introducción

Los robots están listos para llenar un número cada vez mayor de los roles en la sociedad actual, desde la automatización de las industrias hasta aplicaciones de servicio para atención médica y entretenimiento. Inicialmente los robots eran utilizados para tareas repetitivas donde la guía del humano es dada a priori, Ahora participan en tareas y actividades cada vez más complejas y menos estructuradas, incluyendo la interacción con personas que le ayudan a completar las tareas. Esta complejidad ha llevado a la aparición de una nueva disciplina totalmente nueva llamada Interacción Humano-Robot (HRI), el estudio de cómo los seres humanos interactúan con los robots y cual es la mejor manera de diseñar e implementar sistemas de robóticos capaces de realizar tareas interactivas en entornos humanos. El objetivo fundamental de HRI es desarrollar los principios y algoritmos para sistemas de robóticos que los hagan capaces de interactuar de forma directa, segura y eficaz con los seres humanos. Muchas facetas de la investigación en HRI se relacionan y es posible sacar de ideas y principios de la psicología, la comunicación, la antropología, la filosofía y la ética.



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



Definición

La interacción humano-robot (HRI) es el estudio interdisciplinario de la dinámica de interacción entre humanos y robots. Investigadores y profesionales especializados en HRI provienen de una variedad de campos, incluyendo la ingeniería (eléctrica, mecánica, industrial y diseño), la informática (interacción hombre-máquina, la inteligencia artificial, la robótica, la comprensión del lenguaje natural, y la visión por computador), ciencias sociales (psicología, ciencia cognitiva, comunicaciones, la antropología, y factores humanos), y humanidades (ética y la filosofía).

Desafíos de investigación

El estudio de HRI tiene una amplia variedad de retos, algunos de ellos de naturaleza básica de investigación como la exploración de los conceptos generales de HRI y otros de naturaleza específica del dominio que tratan de la utilización directa de los sistemas de robóticos que interactúan con los humanos en contextos particulares. Los principales desafíos de investigación a vista general dentro de HRI son: detección y percepción multimodal; diseño y factores humanos; robots sociales, de servicio y de asistencia; robótica educativa.

Percepción multimodal

La percepción en tiempo real y la disminución de la incertidumbre en la detección son algunos de los desafíos más perdurables de la robótica. Para HRI, los retos de percepción son particularmente complejos, debido a la necesidad de percibir, comprender y reaccionar a la actividad humana en tiempo real.

La diversidad de sensores para la interacción humana es mucho más grande que la mayoría de los otros dominios robóticos en uso hoy en día. Las entradas para HRI incluyen la visión y el habla ambos con grandes retos en el procesamiento de los datos en tiempo real. Sistemas de visión por computador deben procesar datos tales como la expresión facial y los gestos y deben ser capaces de manejar una amplia gama de posibles entradas y situaciones. Del mismo modo, la comprensión del lenguaje y sistemas de diálogo entre los usuarios humanos y robots siguen siendo un desafío de la investigación abierta. Más difícil aún es comprender la conexión entre los datos visuales y lingüísticos y la combinación de estos para mejorar la detección.



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



Diseño y factores humanos

El diseño del robot, en particular lo relacionado con los factores humanos son un aspecto clave de HRI. La investigación en estas áreas se basa en una investigación similar en la interacción humano-maquina(HCI), pero cuenta con una serie de diferencias significativas relacionadas con la física del robot. el aspecto, la forma y el nivel de antropomorfismo y la simplicidad o complejidad del diseño son algunas de las áreas clave de investigación que se están estudiando.

Robots sociales, de servicio y de asistencia

Sistemas robóticos de servicio y asistencia estan incluidos en el amplio espectro de aplicaciones para HRI tales como auxiliares de oficina, asistentes para rehabilitación médica y robots educativos. Esta área integra la investigación básica de HRI con los dominios del mundo real que requieran algún servicio o función auxiliar. El estudio de los robots sociales se centra en la interacción social y por lo tanto es un subconjunto propio de los problemas estudiados bajo HRI.

La robótica de asistencia no ha sido definido formalmente aún pero un robot de asistencia es aquel que brinda ayuda o apoyo a un usuario humano. La investigación en robótica de asistencia incluye los robots de rehabilitación, sillas de ruedas y otros robots ayudantes de movilidad, robots de compañía, brazos manipuladores para personas con discapacidad física. Estos robots están diseñados para su uso en una serie de entornos como escuelas, hospitales y hogares. En el pasado, la robótica asistencial (AR) se ha referido en gran medida a los robots desarrollados para ayudar a las personas a través de la interacción física. Esta definición se ha ampliado considerablemente en los últimos años, en respuesta al creciente campo de la AR en el que los robots proveen ayuda sin necesidad de contacto.

Robótica educativa

La robótica ha demostrado ser una poderosa herramienta para el aprendizaje, no sólo como un tema de estudio, sino también para otros aspectos más generales de la ciencia, la tecnología, la ingeniería y (STEM) matemáticas. Un aspecto central de la educación STEM es la resolución de problemas y los robots sirven como medio para la enseñanza de habilidades de resolución de problemas en entornos grupales. Basado en el éxito de la robótica se han cursos en todo el mundo, ahora hay un movimiento activo para desarrollar hardware y software para robots al servicio de la educación. La robótica se está convirtiendo en una herramienta importante para la enseñanza de la informática y la ingeniería desde los niveles primarios hasta el nivel universitario.



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



Machine Learning

En esta sección se da una idea general de *Machine Learning* basado en el concepto presentado en [8]

Estamos entrando en la era de *Big Data*. Por ejemplo, hay cerca de 1 billón de páginas web; una hora de vídeo es subido a *YouTube* cada segundo, lo que suma 10 años de contenido por día; los genomas de miles de personas, cada una de los cuales tiene una longitud de $3,8 \times 10^9$ pares base y han sido secuenciado por varios laboratorios; *Walmart* se ocupa de más de un millón de transacciones por hora y tiene bases de datos con más de 2,5 petabytes ($2,5 \times 10^{15}$) de información; etc.

Esta avalancha de datos hace necesario métodos automáticos de análisis de datos, que es lo que ofrece *Machine Learning*. En particular, definimos el *Machine Learning* o Aprendizaje de Máquina como un conjunto de métodos que pueden detectar automáticamente patrones en los datos y luego usarlos para predecir datos futuros o para realizar otro tipo de decisiones bajo incertidumbre.

El enfoque probabilístico de *Machine Learning* está estrechamente relacionado con el ámbito de la estadística, pero difiere ligeramente en función de su importancia y de la terminología. En *Machine Learning*, la incertidumbre viene en muchas formas: ¿Cuál es la mejor predicción sobre el futuro dado algunos datos del pasado? ¿Cuál es el mejor modelo para explicar algunos datos? ¿Qué medición se debería realizar en el futuro? etc. *Machine Learning* consiste en una amplia variedad de modelos probabilísticos, adecuados para una amplia variedad de datos y tareas. También una amplia variedad de algoritmos para el aprendizaje y el uso de este tipo de modelos.

Machine Learning suele dividirse en dos tipos principales. En el enfoque predictivo o supervisado, el objetivo es aprender un mapeo de las entradas x a las salidas y , dado un conjunto de etiquetas de entrada-salida $D = \{(x_i, y_i)\}_{i=1}^N$. Donde D se denomina el conjunto de entrenamiento, y N es el número de ejemplos de entrenamiento.

El segundo tipo de *Machine Learning* es el enfoque de aprendizaje descriptivo o sin supervisión. Aquí sólo tenemos, $D = \{x_i\}_{i=1}^N$ y el objetivo es encontrar patrones interesantes en los datos. Esto a veces se llama descubrir conocimiento. Este es un problema mucho menos definido, ya que no se especifica qué tipo de patrones hay que buscar, y no hay una métrica de error evidente (a diferencia de aprendizaje supervisado, donde podemos comparar nuestra predicción y para un x dado contra la etiqueta).



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



Hay un tercer tipo de *Machine Learning*, conocido como aprendizaje por refuerzo. Es útil para aprender cómo actuar o comportarse cuando se dan señales ocasionales de recompensa o de castigo. (Por ejemplo, considere cómo un bebé aprende a caminar.)

Estado del Arte

Dentro de los sistemas de HRI, la generación de gestos visuales del habla o *visual speech* toma gran relevancia debido a la necesidad de la naturalidad en la comunicación oral de un robot hacia los humanos. En este sentido, en los últimos años se han presentado propuestas que buscan la animación automática de los gestos que se producen con el habla. A lo largo del trabajo investigativo que se ha realizado, se han construido métodos basados en la física, también se ha utilizado la selección de fotogramas claves y más recientemente se han incluido métodos de *Machine Learning*.

En los métodos basados en la física, la solución parte en las leyes de la física para determinar el movimiento de la boca, partiendo de algunas condiciones iniciales. Esta técnica requiere modelar los músculos faciales subyacentes y la piel, fue demostrada eficazmente por [9]. En este caso parten de una malla facial estructurada, donde desarrollaron algoritmos que construyen automáticamente modelos funcionales de la cabeza a partir de la toma de información de personas por medio de scanner-laser, sensores de profundidad y reflectancia. Estos algoritmos contraen los músculos en posiciones correctas dentro de un modelo de piel dinámica y los arraigan en una estructura estimada de cráneo con una mandíbula articulada.

En cuanto a los métodos de *key-framing* o fotogramas clave, se especifica alguna secuencia de imágenes (normalmente un fotograma por fonema) e interpolación entre ellas para generar los fotogramas intermedios. Un ejemplo de esto lo aplica LeGoff y Benoit en [10], ellos presentan un sintetizador audiovisual del habla donde realizan un modelo parametrizado en 3-D de la cara (8 parámetros), los valores objetivo de dichos parámetros son asignados dependiendo de cada visema (*visual phoneme*) en pruebas realizadas en un hablante humano. Como resultado se obtienen trayectorias para cada uno de los parámetros del modelo de la cara por medio de funciones de dominancia asociadas a cada parámetro y a cada fonema.

Por último, los métodos de *machine learning* son una nueva clase de herramientas de animación que se forman a partir de datos guardados y luego se utilizan para sintetizar nuevos movimientos. Un ejemplo es el mostrado por Brand en [11], en este caso



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



se logra modelar el comportamiento facial como un Modelo Oculto de Markov o HMM (por sus siglas del inglés, *Hidden Markov Model*) porque todo el contexto necesario para hacer la estimación se puede resumir en un vector de probabilidades de estado actual, y oculto porque no se puede observar los estados, lo que indica que se deben estimar los parámetros del proceso de Markov dados los datos de entrenamiento y una especificación previa de la máquina de estados del modelo.

Ezzat et al. en [12] muestra otro método utilizando herramientas de *Machine Learning*, se busca aprender un modelo de habla animado a partir de información pre-procesada y tomada de un habla predeterminada de una persona y capaz de sintetizar la boca pronunciando enunciados completamente nuevos que no fueron grabados en el vídeo original. Se logra una técnica de síntesis basada en regularización que se entrena automáticamente a partir del habla de un video, y que es capaz de sintetizar trayectorias en el espacio del modelo morfo-métrico correspondiente a cualquier habla deseado.

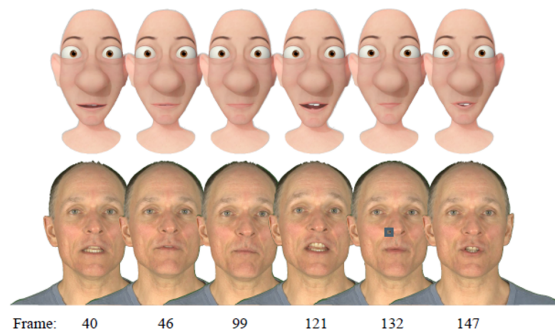


Figura 2: Ejemplo de la secuencia animada generada (fila superior) y sus respectivos fotogramas de un vídeo (fila inferior) [13]

Ahora, la empresa Disney y sus investigadores han realizado desarrollos importantes en este tema ya que sus producciones cinematográficas animadas en 3-D requieren de la animación del habla en sus personajes. Es así como Taylor et al. en [13] presentan un nuevo método que crea una unidad dinámica (visema dinámico) de habla visual que puede generar animación visual realista del habla. El movimiento de los articuladores visuales del habla se utiliza para generar gestos de habla visual discretos y dinámicos. Estos gestos se agrupan (concatenan), proporcionando un conjunto finito de movimientos que describen el habla visual. El sistema aprende nuevas unidades dinámicas (visema dinámico) a partir porciones de video de los datos de entrenamiento, los visemas dinámicos capturan naturalmente la coarticulación visual y la inherente asin-

cronía entre el habla visual y acústica, los datos de entrenamiento también se usan para aprender un mapeo desde fonemas a visemas dinámicos para crear cualquier animación de texto a voz. La figura 2 muestra un ejemplo de la secuencia animada generada y sus respectivos fotogramas de un vídeo.

Luego de esto, Taylor et al. en [14] describen un experimento diseñado para medir el efecto de la tasa de habla o velocidad variable en audio y habla visual, se realiza la comparación de secuencias de fonemas y visemas dinámicos que aparecen en las mismas frases habladas a diferentes velocidades, por medio de este experimento logran demostrar que al variar la velocidad del habla, los sonidos que se producen se modifican, al igual que la posición y la apariencia de los labios, dientes, lengua y otros articuladores visibles. Este es un gran problema ya que los sistemas anteriormente mencionados fueron entrenados para cierta velocidad, por lo que si se quisiera acelerar el dialecto se debería entrenar nuevamente o generar tantos modelos como velocidades de habla se requieran.

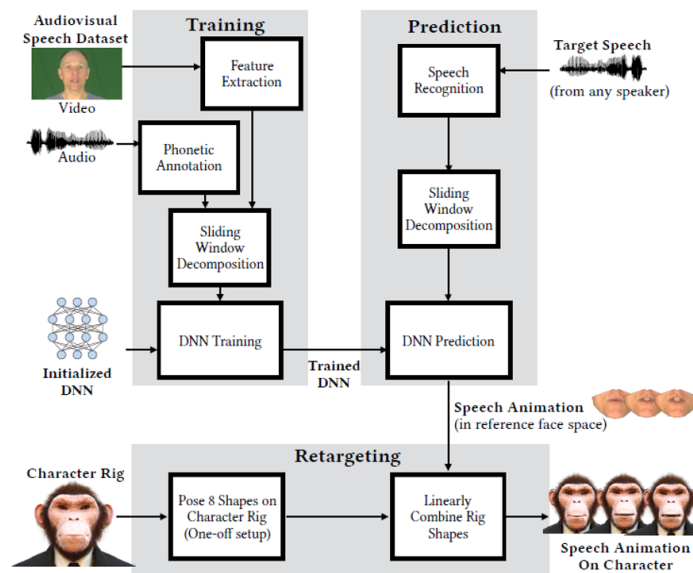


Figura 3: Vista previa del sistema implementado por Taylor et al. en [15]

Presentando una solución a esto, la empresa Disney junto con Taylor et al. en [15] muestran un método de aprendizaje profundo o Deep Learning que logra generar automáticamente una animación de voz de aspecto natural que se sincroniza con el

discurso de entrada. El sistema muestra ciertas propiedades que lo hacen muy robusto: se ejecuta en tiempo real, requiere un mínimo ajuste de parámetros, se genera bien a nuevas secuencias de voz de entrada y se edita fácilmente para crear discurso estilizado y emocional. La figura 3 muestra una vista previa del sistema, en este caso se entrena una red neuronal profunda o DNN (por sus siglas en inglés *Deep Neural Network*) a partir de la información pre-procesada y extraída de los datos de entrenamiento.

Metodología

Este proyecto se basa en la experimentación, la lógica empírica, la observación de fenómenos y su análisis probabilístico. El reto de investigación se centra en la construcción del modelo computacional para la sincronización de los gestos faciales y la voz del robot de manera que favorezca la naturalidad en el contexto de interacción Humano-Robot como se ilustra en la Figura 4.

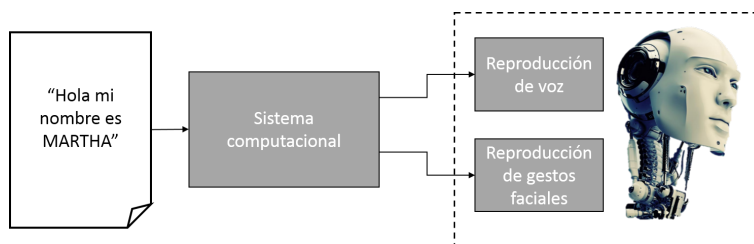


Figura 4: Arquitectura propuesta

La solución del problema se abordará por medio del método de aprendizaje de Máquina o *Machine Learning* y se seguirá la metodología mostrada en la Figura 5.

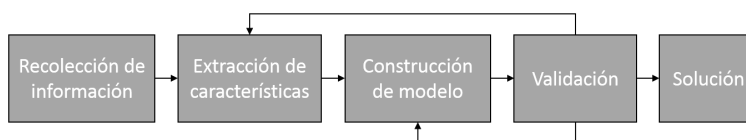


Figura 5: Diagrama de bloques de la metodología

Inicialmente se conforma un banco de información con datos de la gesticulación facial de personas mientras leen textos predeterminados. Dichos datos provienen de un proceso de visión artificial que por medio de un dispositivo de captura de imágenes extrae la

posición de puntos de interés de la cara de la persona, cada gesto capturado por este sistema es asociado al fonema correspondiente del texto.

Luego toda esta información es procesada con el fin de reducir su dimensionalidad, es decir extraer solamente la información relevante la cual permite la construcción de un modelo implementable.

Se construye un clasificador basado en aprendizaje de máquina el cual basado en las demostraciones de las personas crea un modelo que permita la reproducción simultanea de gestos y sonidos al robot.

La validación consiste en una medida cuantitativa probabilística de la coherencia de las predicciones del modelo construido con un conjunto de datos desconocidos. Si la validación revela problemas con el modelo se realizan diferentes iteraciones ajustando la técnica de aprendizaje o el proceso de extracción de características.

Para el proceso de evaluación se usarán dos tipos de criterios: uno cuantitativo para evaluar la exactitud del sistema de aprendizaje de maquina y uno cualitativo por medio de una prueba de experiencia de usuario para evaluar la naturalidad del sistema en el contexto de interacción Humano-Robot.

Herramientas de Hardware

- *SISTEMA DE PROCESAMIENTO*: Computador para la implementación de los algoritmos de procesamiento, entrenamiento y prueba.
- *PROYECTOR DE IMÁGENES*: VideoBeam miniatura para la proyección de una animación articulada de la cara del robot.
- *DISPOSITIVO DE CAPTURA DE IMÁGENES*: Cámara *RGB* o *RGBD* para la detección de la gesticulación de las personas y la construcción del banco de datos.
- *DISPOSITIVO DE REPRODUCCIÓN DE AUDIO*: Bocina convencional para la emisión de la voz del robot.

Herramientas de Software

- *MATLAB*: Herramienta de software matemático con IDE (Entorno integrado de desarrollo) y lenguaje de programación propio que permite el prototipado y validación de algoritmos de forma rápida y sencilla.
- *AUTODESK MAYA*: Software de animación, modelado, simulación y renderizado. Usado para crear el modelo articulado de la cara del robot.



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



- *ROS - Robotics Operating System: Framework* para desarrollo de software para robots.

Resultados esperados

Los resultados de esta propuesta se relacionan con el proyecto global MARTHA. Cuyo objetivo en términos de interacción Humano-Robot es: *Incluir herramientas de hardware, controladores y algoritmos de comunicación bidireccional entre el robot y los humanos de manera multi-modal (visual, sonora, física, etc) que permitan al robot interactuar de manera natural y eficiente con su entorno..* De acuerdo con lo anterior el resultado general que apunta este proyecto es un prototipo funcional de una cabeza robótica que pueda reproducir discursos escritos en forma de voz y gestos faciales con las siguientes características:

- En hardware se espera un prototipo de la cabeza de un robot de servicio con la capacidad de: emisión de sonidos, movimientos en el cuello de dos grados de libertad y gesticulación faciales con al menos 10 grados de libertad.
- En términos de software se espera un modelo articulado y renderizado de la cara del robot y algoritmos originales para la extracción de características y reproducción multi-modal de discursos escritos.

Además se espera la formación en nivel de maestría de un asistente de investigación en el programa de Maestría en Ingeniería Electrónica con énfasis en robótica de la Universidad Santo Tomás.

Productos esperados

Los productos proyectados para este proyecto de acuerdo a la tipología de COL-CIENCIAS se muestran en la tabla 1

Contribución del proyecto al cumplimiento de la misión institucional

El proyecto planteado trabaja sobre las funciones sustantivas investigación y proyección social de la Universidad debido a que sus productos se orientan a la generación de nuevo conocimiento y su objetivo general apunta hacia el mejoramiento de la calidad de vida de la sociedad en general, acorde con los proyectos generales, la misión y visión



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



Cuadro 1: Productos Esperados

PRODUCTO (según tipología COLCIENCIAS)	DESCRIPCIÓN
Productos resultados de actividades de generación de nuevo conocimiento: Artículos de investigación A1, A2, B y C	Un artículo de investigación en revista indexada en ISI o SCOPUS
Productos de resultados de actividades de desarrollo e innovación: Productos tecnológicos certificados o validados	Registro de software de la aplicación desarrollada para la traducción de texto a discurso visual y oral
Productos de actividades relacionados con formación de recurso humano para la CTEI: Proyectos de investigación y desarrollo	Proyecto ejecutado por el Grupo de Estudio y Desarrollo en Robótica GED
Productos de actividades relacionados con formación de recurso humano para la CTEI: Trabajo de grado de maestría	Trabajo de grado de maestría del asistente de investigación vinculado al proyecto (El producto estará sujeto a la continuidad del asistente de investigación en futuros proyectos del grupo)

de nuestro grupo de investigación en robótica

Así mismo la propuesta enmarcada en la robótica de servicio se articula a las líneas de acción del Plan Integral Multicampus PIM 2016-2027 y a las acciones del Plan General de Desarrollo PGD 2016-2019 de la siguiente manera:

Basado en [16] principalmente el proyecto contribuye a la línea de acción **Proyección social e investigación pertinentes** con los siguientes subobjetivos y sus correspondientes acciones:

- *Identificar los campos de acción a nivel nacional y regional de la USTA:* Los robots de servicio trabajan sobre las situaciones cotidianas de las personas por esta razón son una muy buena oportunidad para generar soluciones desde la ingeniería a múltiples problemas en diferentes campos desde el hogar hasta hospitales e industrias.
- *Garantizar que los programas y proyectos de proyección social se articulen con las demás funciones sustantivas, en concordancia con los campos de acción definidos:* El proyecto propone una investigación con impacto social, debido a que la interacción natural entre robots y humanos permitirán acciones cooperativas entre estos



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



en diferentes ámbitos. Entre estos tenemos asistencia médica, búsqueda y rescate de víctimas, transporte de material peligroso y cuidado y monitoreo de adultos mayores entre otros.

- *Incrementar la producción investigativa con impacto regional, nacional e internacional:* Los productos del proyecto apuntan a la generación de nuevo conocimiento con publicación en revista internacional de alto impacto y al desarrollo tecnológico con la producción de prototipos y software.
- *Aumentar de manera progresiva y sostenible la cooperación académica con otras instituciones nacionales e internacionales:* El proyecto está alineado a la filosofía de la red mundial de investigación *Federación RoboCup* y cuenta con la cooperación académica y científica voluntaria de *RoboCup@Home League*.

Además el proyecto genera un desarrollo tecnológico que sirve para nuevas investigaciones y para la enseñanza de la robótica y la inteligencia artificial. Esta herramienta diseñada para el laboratorio de robótica de la Universidad contribuye a la línea de acción **Capacidad y gestión institucional que logran la efectividad multicampus**, más precisamente a las acciones del subobjetivo fortalecer la infraestructura física, tecnológica y de servicios con mínimos comunes de la USTA.



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



Presupuesto

Recurso solicitado FODEIN		
Concepto	Descripción	Monto
Personal científico	Carlos Saith Rodríguez Rojas	60h/mes, \$17'943.750
	Eyberth Rolando Rojas Martínez	40h/mes, \$14'947.500
Auxilio a investigadores	No aplica	\$0
Asistentes de investigación	Reconocimiento económico al estudiante de maestría	\$7'000.000
Equipos	Se usarán los equipos disponibles en el laboratorio de robótica	\$0
Software	Se usará software libre y versiones educacionales sin costo	\$0
Materiales	Proyector de imágenes portátil Máscara acrílica semi-transparente Cámara 3D Servomotores Acoples mecánicos	\$8'500.000
Papelería	No aplica	\$0
Fotocopias	No aplica	\$0
Salidas de campo	No aplica	\$0
Material bibliográfico	Se usarán los recursos físicos y electrónicos de la biblioteca	\$0
Publicaciones	No aplica	\$0
Servicios técnicos	No aplica	\$0
Movilidad académica	Participación en conferencia internacional	\$3'000.000
Organización de eventos	No aplica	\$0
	Total	\$50'891.250
Contrapartida externa		
Para proyectos en cooperación y alianza estratégica		
Institución	Descripción	Monto
	Total	



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



Institución de Educación Superior sujeta a la inspección y vigilancia del Ministerio de Educación Nacional - SNIES 1704

Cuadro 2: My caption

Actividad	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8	Mes 9
Revisión y caracterización de técnicas del estado del arte									
Construcción de la cabeza para el robot social									
Diseño y desarrollo de la cara virtual articulada									
Diseño del experimento para recolección de datos y construcción de modelo computacional									
Realización de informe de avance									
Montaje del experimento para recolección de datos									
Análisis de datos y entrenamiento de modelo computacional									
Integración de sistema de reproducción de voz									
Diseño y aplicación de pruebas de experiencia de usuario									
Escritura de artículo para conferencia internacional									
Realización de informe final del proyecto									



Cronograma

Bibliografía

- [1] “Robocup federation.” [Online]. Available: <http://www.robocup.org/>
- [2] T. Kanda, *Natural Human-Robot Interaction*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 2–2.
- [3] G. I. of Technology, C. M. University, R. T. Consortium, U. of Pennsylvania, U. of Southern California, S. University, U. of California-Berkeley, U. of Washington, and M. I. of Technology, *A Roadmap for U.S. Robotics: From Internet to Robotics*. Georgia Institute of Technology, 2013. [Online]. Available: <https://books.google.com.co/books?id=KPhQngEACAAJ>
- [4] S. c/o euRobotics AISBL, *Robotics 2020 Multi-Annual Roadmap For Robotics in Europe*. SPARC, 2015. [Online]. Available: https://www.eu-robotics.net/cms/upload/downloads/ppp-documents/Multi-Annual.Roadmap2020_ICT-24_Rev_B_full.pdf
- [5] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, and E. Osawa, “Robocup: The robot world cup initiative,” in *Proceedings of the First International Conference on Autonomous Agents*, ser. AGENTS ’97. New York, NY, USA: ACM, 1997, pp. 340–347. [Online]. Available: <http://doi.acm.org/10.1145/267658.267738>
- [6] J. Stuckler, M. Schwarz, and S. Behnke, “Mobile manipulation, tool use, and intuitive interaction for cognitive service robot cosero,” *Frontiers in Robotics and AI*, vol. 3, p. 58, 2016. [Online]. Available: <http://journal.frontiersin.org/article/10.3389/frobt.2016.00058>
- [7] D. Feil-Seifer and M. J. Matarić, *Encyclopedia of Complexity and Systems Science*. New York, NY: Springer New York, 2009, ch. Human RobotHuman–robot interaction (HRI) InteractionInteraction human robot, pp. 4643–4659. [Online]. Available: http://dx.doi.org/10.1007/978-0-387-30440-3_274
- [8] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.
- [9] Y. Lee, D. Terzopoulos, and K. Waters, “Realistic modeling for facial animation,” in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. ACM, 1995, pp. 55–62.



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6



- [10] B. Le Goff and C. Benoît, “A text-to-audiovisual-speech synthesizer for french,” in *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, vol. 4. IEEE, 1996, pp. 2163–2166.
- [11] M. Brand, “Voice puppetry,” in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1999, pp. 21–28.
- [12] T. Ezzat, G. Geiger, and T. Poggio, “Trainable videorealistic speech animation,” *ACM Trans. Graph.*, vol. 21, no. 3, pp. 388–398, Jul. 2002.
- [13] S. L. Taylor, M. Mahler, B.-J. Theobald, and I. Matthews, “Dynamic units of visual speech,” in *Proceedings of the 11th ACM SIGGRAPH/Eurographics conference on Computer Animation*. Eurographics Association, 2012, pp. 275–284.
- [14] S. Taylor, B.-J. Theobald, and I. Matthews, “The effect of speaking rate on audio and visual speech,” in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 3037–3041.
- [15] S. Taylor, T. Kim, Y. Yue, M. Mahler, J. Krahe, A. G. Rodriguez, J. Hodgins, and I. Matthews, “A deep learning approach for generalized speech animation,” *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, p. 93, 2017.
- [16] U. S. Tomás, “Documento síntesis plan integral multicampus 2016-2027 y plan general de desarrollo 2016-2019,” 2016. [Online]. Available: <http://planeacion.usta.edu.co/images/documentos/PIM.DOCUMENTO.pdf>

Posibles evaluadores

- **Fernando Enrique Lozano Martínez PhD**, Profesor Asociado de la Universidad de los Andes, flozano@uniandes.edu.co
- **Fernando de la Rosa Rosero PhD**, Profesor Asociado de la Universidad de los Andes, fde@uniandes.edu.co
- **Henry Carrillo Lindado PhD**, Profesor Asistente de la Universidad Sergio Arboleda, henry.carrillo@usa.edu.co



BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6

