

INTERFACE NATURAL EMOTIVA NIÑO - ROBOT.

Néstor Balich^{#1}, Alejandro Cena², Alejandro Fransoy^{#3}

Laboratorio de Robótica Física LRF /
Centro de Altos Estudios en Tecnología Informática CAETI /
Facultad de Tecnología Informática /
Universidad Abierta Interamericana

Av. Montes de Oca 745 - (+54 11) 4301-5323 - C.A.B.A. - República Argentina

<http://lrfuai.blogspot.com.ar/>

¹nestor.balich@uai.edu.ar

²alejandro.cena@uai.edu.ar

³alejandromarcelo.fransoy@alumnos.uai.edu.ar

Resumen

El objetivo de esta investigación es la de desarrollar conocimientos en el área de la interacción natural en ambientes dinámicos, con el objetivo de alcanzar un nivel de precisión similar al que obtenemos con el uso de periféricos tradicionales a través de una HRI (Human Robot Interface) . La interacción natural es una tendencia que permite vincularnos a los dispositivos tecnológicos de una manera más transparente y empática que con el enfoque tradicional. Estamos convencidos que rápidamente este tipo de interacción va a avanzar por sobre el tradicional, como se evidencia en nuevos dispositivos como smarttvs, teléfonos celulares y consola de video juegos.

Para lograr este objetivo nos enfocamos en la utilización del dispositivo Kinect combinado con la librería de reconocimiento del Habla y Síntesis de Voces de Microsoft. En la siguiente investigación, dejamos plasmada nuestra experiencia tanto en laboratorio como en trabajos de campo donde detallamos los inconvenientes que tuvimos y como pudimos sortearlos, junto con la perspectiva de mejora a futuro.

Palabras clave: robot, kinect, reconocimiento de voz, HRI, interface emotiva, interface verbal, reconocimiento en tiempo real.

Contexto

El proyecto está enmarcado en el desarrollo de una unidad robótica que permita la ejecución de tareas variadas en ambientes dinámicos brindando una interface de interacción natural con su contexto. Este proyecto se inició en Abril de 2012 con el diseño y fabricación del robot RobIN dentro del Laboratorio de Robótica Física Aplicada (LRF), del Centro de Altos Estudios en Tecnología Informática (CAETI), dependiente de la Facultad de Tecnología Informática de la Universidad Abierta Interamericana (UAI). Las líneas investigativas se orientaron a diferentes

ramas de la interacción natural con el entorno (humano y físico), siendo las principales el Reconocimiento Visual, Interacción natural Hombre-Máquina, Navegación Inteligente en ambientes dinámicos, Seguimiento de Objetivos y Coordinación multiagente.

El proyecto aquí presentado se enmarca en la línea de La Interacción Natural Hombre-Máquina. Con el objetivo de llegar a un mecanismo de interacción que permita superar las experiencias que proponen el manejo de dispositivo electrónico (robot).

Considerando especialmente la relación costo/beneficio de las tecnologías y materiales utilizados, el aporte a la comunidad y desarrollo

Introducción

En la teoría de la comunicación se encuentran presentes 3 componentes (Emisor - Medio - Receptor), en el caso de la interacción, esta comunicación es del estilo Estimulo-Medio-Respuesta entre el usuario y el dispositivo a controlar, el cual se basa en el mismo principio y plantea como eslabón más débil el Medio. Sobre este punto, el enfoque tradicional ofrece al usuario un periférico que garantiza la sanidad del canal entre el emisor-receptor simplificando la comunicación. Los enfoques más actuales proponen lidiar contra un medio no controlado (o dinámico), en pos de emular a la interacción natural que se realiza entre 2 personas.

¿Cómo medir la calidad de la interacción?

El resultado de la interacción exitosa con la interface se observó empíricamente con el grado de retroalimentación que se obtuvo con el individuo, la cantidad de órdenes, el tiempo involucrado hasta finalizar la interacción, la respuesta emotiva a acciones específicas y la transmisión del conocimiento por observación o por aprendizaje a nivel usuarios, sin intervención de expertos ni manuales.

Centramos esta fase del proyecto y este paper en particular en la interacción verbal mediante ordenes, interacción sonora asociados a la memoria emotiva con eventos de la vida diaria, programas de televisión, películas, animales, dibujos animados y música, todas con temáticas asociada al interés de los niños.

II.1 Interface Natural

La interacción entre el mundo real y virtual siempre ha representado un desafío para quienes se adentren en el ámbito de las ciencias robóticas. En las últimas décadas muchos fueron los avances en ésta rama, considerando aspectos tan difundidos como la interfaz táctil que encontramos en mouse, teclados, pantallas, botones de mandos, como otros más avanzados que van desde la captura de movimientos humanos por medio de sensores visuales sofisticados, kinéticos y hasta reconocimientos de voz en tiempo real.

En el proceso de captación de señales naturales en un mundo digital, es de vital importancia el canal de comunicación que se establece entre el emisor y el receptor. En la actualidad, los sistemas más sofisticados de reconocimiento sonoro permiten operar en niveles bajos de ruido en el canal de forma de conocer de manera precisa, qué tan confiable es la traducción que ha realizado el módulo del sonido captado.

Nuestro trabajo pone énfasis en la interacción natural entre niños y el robot. La manera en que se establecen lazos de comunicación de forma natural por medio de lenguaje de gestos, sonidos y su vínculo con la memoria emotiva.

Cuando un niño asimila que puede relacionarse con un robot a través de un dialogo percibiendo reacciones físicas que puede asociar a sus conocimientos, su confianza va aumentando rápidamente permitiendo reflejar la creatividad y espontaneidad característica en infantes y, en consecuencia, facilitar el aprendizaje.

Habiendo experimentado en interfaces a nivel adultos en el laboratorio, propusimos integrar y avanzar con una HRI basada en la realimentación visual, sonora, kinética dotando de personalidad al robot empleando sonidos, frases y movimientos que permitan mantener un dialogo simple llegando a un total de 30 sonidos imitando animales 20 sonidos mediáticos, 20 frases 4 movimientos y luces, orientada a la interacción con niños.

II.2 Memoria Emotiva

La memoria emotiva es aquella que vincula un recuerdo con una experiencia emotiva. Tiene gran influencia en el aprendizaje y desarrollo de los niños, quienes suelen utilizar constantemente este vínculo entre las emociones, las experiencias vividas y el refuerzo de las redes de memoria. Aplicado a las interfaces naturales implica la búsqueda de un resultado emotivo a través de la interacción con herramientas tecnológicas.

Motivados por ello, decidimos centrar la investigación en la interacción del grupo etario dentro del rango de 5 a 12 años, con nuestro robot ROBIN (Robot de Interface Natural) (Figura 1).



Figura 1: RobIN interactuando en Tecnópolis

Líneas de Investigación, Desarrollo e Innovación

Este proyecto se enmarca dentro de las líneas de investigación de ingeniería de software dentro del WICC (IS) nuevas tecnologías educativas y diseño de interfaces en el marco general y automatización y robótica en la líneas de investigación del CAETI.

El 100% de estos desarrollos son articulados en forma directa como herramienta educativa de carácter experimental, didáctico y motivador de la materia Robótica que se dicta de forma combinada en el laboratorio de computación de la universidad y el LRF-CAETI.

Arquitectura RobIN

RobIN se compone de un chasis que contiene una interface de potencia para actuadores y sensores basada en un PIC16F873, alimentación a batería de gel de 12 Volt, un soporte superior para el sensor kinect y una base plana que sostiene una notebook como controlador principal.

La arquitectura de la solución se basa en el proyecto SARA (de las siglas de Software Autónomo de Robótica Aplicada) desarrollado en el laboratorio. Esta arquitectura propone la separación de las responsabilidades en 5 capas. Desde el nivel más bajo al de mayor abstracción sería:

- Capa Hardware
- Capa Firmware
- Capa Lógica
- Capa Funcional
- Capa Ejecución

Capa de Hardware se compone de un diseño de open hardware ya que controla al robot .

Capa Firmware es la capa que contiene el software de bajo nivel con conocimiento del manejo de los diferentes componentes de hardware, brindando una interfaz unificada para la comunicación con las capas superiores. Esto nos permite Modificar el diseño de la placa de manera transparente hacia las capas superiores.

Capa Lógica tiene la responsabilidad de interpretar las instrucciones a realizar traduciéndolas con la versión de firmware propia del robot. Esto nos permite independizar las capas superiores de la comunicación con el Firmware, permitiéndonos la reutilización. Ejemplos: (driver de Kinect, driver de Reconocimiento del Habla, Sintetizador de Voz, Navegación, etc.)

Capa Funcional tiene la responsabilidad de contener el repositorio de funcionalidades que van a ser usadas en componentes de más alto nivel. Estas pueden combinarse entre sí para lograr comportamientos más complejos. La idea de esta capa es brindar componentes con la suficiente complejidad inherente como para ser utilizados en las capas superiores. Entre los componentes se encuentran diferentes tipos de navegación, seguimiento de objetivos, manejo de repositorios de gramática para la detección del Habla (Figura 2), etc.

```

1 </grammar>
2 version="1.0" xml:lang="es-ES"
3 root="rootRule" tag-format="semantics/1.0-literals"
4 xmlns="http://www.w3.org/2001/06/grammar">
5 <rule id="rootRule">
6 <one-of>
7 <item>
8 <tag>COMANDO</tag>
9 <one-of>
10 <item> Frase 1 </item>
11 <item> Frase N </item>
12 </one-of>
13 </item>
14 <....>
233 <item>
234 <tag>DECIR_FECHA</tag>
235 <one-of>
236 <item> que dia es hoy </item>
237 <item> que fecha es hoy </item>
238 </one-of>
239 </item>
240 </one-of>
241 </rule>
242 </grammar>

```

Figura 2: Ejemplo de comandos gramaticales en XML

Capa de Ejecución es la que se encarga de controlar la ejecución de la rutina de alto nivel, que combine diferentes unidades funcionales para la concreción de un objetivo. Que en este caso es la interface de interacción natural.

Resultados y Objetivos

- Trabajo de campo

Habiendo experimentado en laboratorio quisimos llevar la experiencia un paso más, aprovechando el espacio cedido por el CESSI en Tecnopolis en donde se decidió poner a prueba la interface y la potencialidad en la transferencia de conocimiento de su uso en el grupo objetivo (principalmente niños).

Esto nos permitió a probar la interface en función de la cantidad de frases, respuestas y sonidos por observación directa de más de 500 personas en grupos de hasta 10.

- Problemas.

Las condiciones encontradas, diferentes a las experimentadas en el laboratorio, nos llevó a realizar diferentes modificaciones. Dichos se detallan a continuación:

Ruido Ambiente: Primer problema que experimentamos en el reconocimiento del habla, dado factores ambientales, cantidad de personas y otros robots.

Direccionalidad de la Captura. En la configuración inicial capturábamos el sonido con los dos micrófonos disponibles en el Kinect. En el laboratorio nos permitió ubicar la fuente de sonido pero al estar expuestos a un gran ruido ambiente excediendo los umbrales de funcionamiento hizo que dicha funcionalidad presentara un comportamiento errático.

Umbral de Confianza. Este es el valor que nos permite medir la probabilidad de acierto sobre la palabra reconocida y asociarlas a acciones. Los altos valores obtenidos en condiciones de laboratorio, se vieron drásticamente disminuidos al exponernos a los altos niveles de ruido ambiente.

Solución. Todos estos factores detallados anteriormente tienen de común denominador el ruido ambiente. Para poder sortear estos inconvenientes era necesario aumentar la nitidez del sonido a capturar, se decidió aumentar la direccionalidad de la captura utilizando un micrófono direccional, con el que logramos disminuir la influencia del ruido ambiente manteniendo el nivel de interacción.

Otros Problemas. A medida que se iban sucediendo las interacciones con el Robot, identificamos problemas inherentes a la propia naturaleza de los chicos. Evidenciamos problemas con la **Fonética**, ya que la librería presentaba dificultades en identificar las palabras debido principalmente a **la rapidez al hablar** propio de la ansiedad y los nervios, dificultad en la **pronunciación** de las palabras, esto es más notorio en los chicos más pequeños. Y por último una limitación de la librería que ya se evidenciaba en condiciones de laboratorio, **la internacionalización**, ya que la librería está disponible para español tradicional, con lo cual teníamos algunas dificultades en la pronunciación en palabras, en especial con la “ll” siendo que tendemos localmente a pronunciarlo como una “y”.

Puntos de Mejora. Teniendo en cuenta la experiencia de campo y la futura integración de esta investigación en un proyecto más amplio y contenedor, detectamos importante realizar algunas mejoras en el área de la direccionalidad de la detección del sonido. Para esto creemos que sería importante equipar al robot con un

micrófono direccional y combinarlo con la funcionalidad del Kinect de detección de individuos, para primero identificar a la persona con quien realizara la interacción y luego direccionar los micrófonos hacia él. Esto combinado con filtros electrónicos y de software estimamos llegar a una interacción más cercana a la planteada en condiciones de laboratorio.

Formación de Recursos Humanos

El equipo de trabajo del LRF-UAI cuentan con 2 Profesores Universitarios 1 Educador a nivel Secundario y 6 alumnos de la carrera de ingeniería en sistemas de la UAI, como auxiliares de investigación, uno auxiliar de la materia Robótica de 3° año de dicha carrera.

En esta fase del proyecto de interfaz emotiva, han trabajado al 100% un total de 4 alumnos universitarios, 1 alumno secundario, en 2012 Alejandro Cena y Laura Carreras y Franco Balich continuado en 2013/2014 por Alejandro Cena , Alejandro Fransoy (alumno de ingeniería en sistemas) con la dirección y participación del Ing. Néstor Balich.

Durante la cursada de la materia Paradigmas Tecnológicos II (Robótica) 2012/2013 varios grupos de alumnos han realizado testeos y sugerencias sobre este desarrollo. Muchos de los cuales continúan en proyectos del LRF extendiéndose en prácticas reales, investigación y desarrollo.

2011/ 2012 / 2013 Fabricación de robot RobIN Colaboradores Franco Balich, Carlos Saavedra profesor Néstor Balich

2012 Primera versión verbal de SARA, Tecnopolis CIITI 2012 Alumnos Alejandro Cena, Laura Carrera profesor Néstor Balich

2013 Interface gestual en python InnovatiBA, CIITI 2013 Rosario/ Bs As Alumnos Alejandro Cena profesor Néstor Balich

2014 Segunda versión SARA Alumnos Alejandro Cena, Marcelo Fransoy

2014 Incorporación auxiliar de materia Alejandro Cena

Agradecimientos

Deseamos agradecer a la Universidad Abierta Interamericana, al Centro de Altos Estudios en Tecnología Informática, Dr Marcelo DeVincenzi, CESSI y Tecnópolis por apoyar el trabajo realizado por alumnos, pasantes e investigadores del LRF permitiéndonos difundir los trabajos de investigación, transferencia tecnológica, formación de recursos humanos y desarrollo de tecnología en el área de robótica educativa .

Referencias

- [1] Cory J. Hayes, Charles R. Crowell, and Laurel D. Riek, "Automatic Processing of Irrelevant Co-Speech Gestures with Human but not Robot Actors", Human-Robot Interaction (HRI), 8th ACM/IEEE International Conference 3-6 Marzo 2013
- [2] Guy Hoffman, Gurit E. Birnbaumy, Keinan Vanunu y, Omri Sass, Harry T. Reis , "Robot Responsiveness to Human Disclosure Affects Social Impression and Appeal" , ACM/IEEE international conference on Human-robot interaction HRI '14 Proceedings of the 2014
- [3] Radford Parker, "Real-Time Kinect Player" , Rumania, Agosto 2012
- [4] Georgios Galatas, Gerasimos Potamianos, and Fillia Makedon, "Audio-Visual Speech Recognition Incorporating Facial Depth Information Captured by the Kinect" , Rumania , Agosto 2012
- [5] Matthias Kronlachner and IOhannes m zmolnig, "The Kinect Sensor as Human-Machine-Interface in Audio-Visual Art Projects", Austria , Septiembre 2012.
- [6] Guy Hoffman, "Embodied Cognition for Autonomous Interactive Robots" Cognitive Science Society 2012
- [7] Min-Joon, Jin-Wook Beak and In-Kwon Lee, "Creating Musical Expression using Kinect" , Noruega, Mayo 2011,.
- [8] Marek P. Michalowski, Selma Sabanovic, Philipp Michel "Roillo: Creating a Social Robot for Playrooms" The 15th IEEE International Symposium on Robot and Human Interactive Communication, Hatfield, UK, Septiembre 6-8, 2006
- [9] Terrence Fong, Laura M. Hiatt, Clayton Kunz, Magda Bugajska "The HumanRobot Interaction Operating System", CMU, 2006
- [10] Soshi Iba, Tesis "Interactive Multi-Modal Robot Programming", The Robotics Institute Carnegie Mellon University, Mayo 3 2004
- [11] J. Schulte, Spontaneous, C. Rosenberg, S. Thrun, "Short-term Interaction with Mobile Robots" School of Computer Science Carnegie Mellon University - IEEE International Conference, Robotics and Automation, 1999