

Propuesta de framework de adaptación de aplicaciones multimedia con base a señales sociales

Carlos Ivan Rivera Parra^{1*}

Resumen

Las aplicaciones multimedia requieren que la interacción con los usuarios sea de forma más natural y poderse retroalimentar de los estados que el usuario expresa, es importante poder incluir las señales sociales que el usuario presenta en el momento de la interacción con una aplicación y de esta forma poder adaptar las aplicaciones multimedia en tiempo real en base a un conocimiento más personal del usuario, por este motivo en este trabajo se presenta el desarrollo de un framework de adaptación de aplicaciones multimedia basado en señales sociales, el cual permitirá integrar frameworks de procesamiento de señales sociales con las aplicaciones multimedia para obtener la detección de las señales emitidas por el usuario y adaptarlas de acuerdo a un aprendizaje hecho del uso de la aplicación.

Abstract

Multimedia applications require interaction with users is more naturally and be able to feed back to the states that expresses the user, it is important to include social signals that the user has at the time of interaction with an application and thus to adapt multimedia applications in real time based on a more personal knowledge of the user, for this reason, in this work the development of a framework for adapting multimedia applications based on social cues, which will integrate frameworks signal processing is presented social with multimedia applications for the detection of the signals emitted by the user and adapted according to a fact learning-use application

Palabras Claves

HCI, SSP, Aplicaciones Multimedia, Señales sociales, Adaptación.

¹ Universidad Javeriana, Bogotá, Colombia

*Contacto: carlosrivera@javeriana.edu.co

Índice

Introducción	1
1 Herramientas existentes de SSP	2
2 Framework propuesto	3
2.1 Frameworks SSP	3
2.2 Identificación de comportamientos	4
2.3 Base de comportamientos	4
2.4 Aprendizaje de reglas	4
2.5 Reglas de adaptación	4
2.6 Interfaz Gráfica	5
2.7 Aplicación Multimedia	5
3 Conclusiones	5
Referencias	5

Introducción

La forma en que los seres humanos interactúan con los objetos de su entorno, la personas a su alrededor y cómo reaccionan a estas interacciones se realiza a través de la comunicación verbal y no verbal, la comunicación no verbal es el principal canal con el cual nos comunicamos con los demás. Las personas no pueden dejar de comunicarse, en cada

momento están mostrando diferentes señales comportamentales no verbales las cuales pueden ser expresiones faciales, gestos corporales, apariencia, etc. También de otra forma las personas reciben las señales que otros expresan, comúnmente de manera inconsciente se interpretan las señales de los otros en forma de emociones, sentimientos, actitudes, intenciones, entre otros [1].

Este tipo de comunicación no verbal se ha convertido en uno de los desafíos más grandes a los que se enfrentan los científicos, el cómo identificarlos y traducirlos en comportamientos, estos comportamientos de las personas son expresiones de sus actitudes frente a diversas situaciones en el entorno y cómo interactúan con ellas, produciendo señales y comportamientos sociales como empatía, desacuerdo, cortesía, entre otros. Estas señales sociales pueden ser obtenidas a través del análisis y combinación de diversas señales de comportamiento (behavioural cues) las cuales pueden ser como se nombraron anteriormente expresiones faciales, expresiones vocales, gestos corporales, entre otras, en la figura 1 se puede observar como una combinación de señales de comportamiento producen una señal social, que en el caso de la figura es desacuerdo o agresividad, lo cual es evidente inclusive si se mostraran solo las siluetas de los personajes involucrados en la interacción [2].

Procesamiento de señales sociales (SSP, por sus siglas en

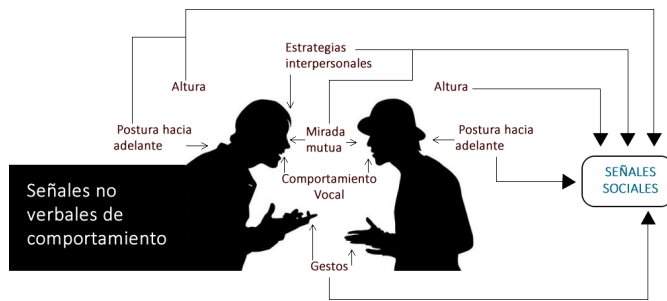


Figura 1. Señales de comportamiento

inglés) es el campo de investigación de dominio tecnológico que busca que los computadores y sistemas tengan la capacidad de realizar un análisis, detección y reconocimiento automático de señales sociales. Este campo involucra la ingeniería y las ciencias humanas y busca analizar desde estas dos perspectivas los comportamientos entre las interacciones hombre-máquina y hombre-hombre, permitiendo interpretar los comportamientos no verbales expresados por las personas en su interacción con otras personas o máquinas en su entorno social inmediato [3].

Como se puede ver, la importancia que las señales sociales tienen en la vida diaria del ser humano es evidente, por eso se hace necesario la forma de poder evaluar automáticamente estos comportamientos sociales de las personas [2], esto se hace posible a través el procesamiento de señales sociales (SSP, por sus siglas en inglés), el cual tiene como objetivo acercar y reducir la brecha de la inteligencia social entre computadoras y humanos [4]. El SSP está interesado en responder las siguientes preguntas [1].

- ¿Es posible detectar señales comportamentales no-verbales en grabaciones de interacciones sociales capturadas con micrófonos, cámaras y otras clases de sensores?
- ¿Es posible inferir señales sociales desde señales comportamentales no-verbales a través de datos capturados con diferentes clases de sensores?
- ¿Es posible sintetizar señales comportamentales no-verbales provocando percepciones sociales deseadas?

En respuesta a estas preguntas y el estudio de SSP se han encontrado diferentes campos de aplicación del procesamiento de señales sociales, los cuales son análisis de interacciones humanas, entrenamiento (coaching), robótica social, e interacciones con ambientes virtuales [4]. Para lograr los desarrollos en estas aplicaciones se han tenido que desarrollar e involucrar diferentes tipos de detección de señales sociales, para esto ha sido necesario el uso de diferentes clases de sensores y que permitan la detección de estas señales, en la mayoría de aplicaciones de SSP se encuentra que cada investigador propone y construye su método de detección y análisis de las señales para hacerlo y aplicarlo a su contexto específico, esto hace que el investigador no pueda enfocarse solo en su

Cuadro 1. Herramientas SSP

HERRAMIENTA	FUNCIÓN	AÑO
iBUG Smile Detectors	Deteccion de sonrisas	2012
Social Signal Interpretation Framework (SSI)	Framework que ofrece herramientas para grabar, analizar y reconocer comportamientos humanos en tiempo real (gestos, mimicas, y discursos emocionales)	2012
Sneaker Diarization Toolkit	Speaker diarization	2012
iBUG TAUD	Deteccion de Action Units	2011
iBUG Gesture Detector	Deteccion de Gestos de las manos	2011
Elckerlvc	comportamientos multimodales verbales y no verbales para Humanos Virtuales	2011

campo de estudio relevante sino que debe preocuparse también de cómo realizar la detección y análisis necesario para su aplicación específica, en respuesta a esto se han desarrollado diversos frameworks los cuales permiten generar un análisis de señales a través de diversos sensores (micrófonos, cámaras, etc...) [5][6][7][8]. Estos frameworks se han podido utilizar en diversas aplicaciones y diferentes proyectos tanto en el ámbito académico como fuera de él [9], a pesar de que estos permiten el uso de diferentes sensores de detección y generan reconocimiento en tiempo real de señales sociales, no existe una forma de empalmarlos fácilmente en diversos sistemas, este proceso requiere de una configuración especial a cada caso de uso, la mayoría de las

aplicaciones en las cuales se han utilizado responden al análisis de las señales e interpretar las señales emocionales de usuarios [10][11], se presentan usos para la detección de desórdenes psicológicos [12].

1. Herramientas existentes de SSP

La red europea de excelencia en SSP (www.SSPNet.eu) tiene una recopilación de diversas herramientas que hacen uso y contribución a SSP, que han sido resultados de múltiples investigaciones hechas en el tema. En el cuadro 1 se listan una muestra de las herramientas analizadas que reporta SSPNet y se da una breve descripción de su función, de este análisis hecho se extrajeron las características de cada una de ellas, como el lenguaje en el cual fueron desarrolladas, la plataforma en la que se implementa, descripción y para algunos casos las entradas y salidas que producían, se puede ver la lista completa y más detalle en la página de SSPNet.

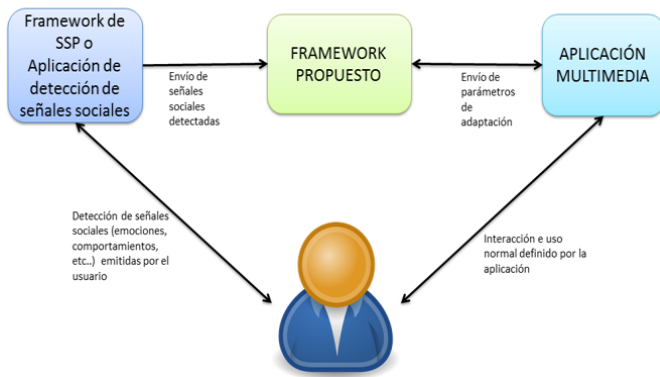


Figura 2. Objetivo del modelo propuesto

2. Framework propuesto

Como se ha mencionado anteriormente las señales sociales y su procesamiento están presentes en cada interacción de las personas y su entorno; SSP ayuda a estudiar y analizar los comportamientos que los seres humanos tienen ante diversas situaciones que se presentan a su alrededor, el poder interpretar y determinar cómo las señales sociales son producidas y en qué momento para una persona en específico, ayuda a identificar los factores que producen este tipo de señal en la persona e identificar aspectos que puedan afectar su relación con el entorno. La interacción hombre maquina/computador (HCI por sus siglas en inglés, human computer interaction) ha diseñado sistemas con un acercamiento enfocado a diseños centrados en computadores, lo cual para aplicaciones y sistemas que sean independientes del contexto funciona correctamente, como lo es el caso de sistemas para reservaciones de tiquetes de aviones, en el cual no es necesario tener interacciones complejas y con altas funciones; el caso de aplicaciones en las cuales se deben interactuar con diferentes sistemas y en ambientes más complejos no funciona correctamente. Por esta razón se ha cambiado del concepto de diseño centrado en computadoras y se ha cambiado a diseño centrado en humanos el cual es basado en modelos de comportamiento humano y requiere información del que se comunica, satisfacer las necesidades y requerimientos de los usuarios.

El diseño centrado en humanos se hace necesario en el diseño y desarrollo de aplicaciones multimedia que hacen uso de diferentes medios para interactuar con el usuario, mostrarle información, y adaptarse en su medida a las características y acciones de este. El desarrollo de aplicaciones con diseño centrado en usuario es un campo de aplicación de SSP debido a los requerimientos que este tipo de diseño tiene sobre el entendimiento de los comportamientos humanos y su comunicación. Este entendimiento de los comportamientos permite generar una interacción natural y no convencional en los humanos y las maquinas/computadores. Uno de los desafíos de los sistemas multimedia es cómo integrar inteligencia en los ambientes y objetos, crear sistemas capaces de entender,

inferir e influenciar el entorno [13].

En las aplicaciones multimedia es importante la interacción del usuario con el sistema, Es necesario que se adecuen para producir una interacción más personal con el usuario y que esta sea de forma natural. Un ejemplo de aplicaciones multimedia es los videojuegos, los cuales muchas veces pueden tener contenido rígidos y estáticos, lo cual puede resultar en que el juego parezca predecible e impersonal, por esta razón es necesario introducir la adaptación de estos sistemas, existen avances en este tema de los cuales se puede ver 2 clasificaciones las off-line y las online, las offline buscan introducir más información del usuario previamente para poder adaptar el sistema y las online se han trabajado más en la adaptación de ambientes de juego o de los personajes no jugadores (NPC por sus siglas en inglés), que son inteligencias artificiales en el juego y se adaptan de acuerdo las decisiones y acciones que el usuario toma durante el juego. Estas adaptaciones no tienen en cuenta el estado actual de las personas y sus comportamientos, por lo tal está abierto el desafío en los juegos de incluir estos comportamientos y la información del jugador en tiempo real para adaptar el juego. Como se puede ver en el caso de os juegos es una oportunidad evidente y esto también es posible llevarlo a diversas aplicaciones multimedia, debido a que varios de las metas de SSP se alinean con las características y objetivos de las aplicaciones multimedia.

El framework propuesto se enfoca en manejar la adaptación de las aplicaciones multimedia, basado en señales sociales (figura 2), las cuales son producidas por diversas herramientas o Frameworks de señales sociales como las vistas en la sección anterior, de acuerdo a las salidas arrojadas por estos Frameworks se adaptaran algunos parámetros de las aplicaciones multimedia y harán que la interacción con el usuario sea personalizada y ajustada de acuerdo a la respuesta de este con el sistema. El framework busca que se puedan integrar varios Frameworks de señales sociales para enriquecer la experiencia y el conocimiento del usuario en pro de una aplicación multimedia que se desee adaptar de acuerdo al usuario. En la figura 3 se muestra el diagrama de componentes del framework propuesto para poder adaptar aplicaciones basadas en señales sociales, a continuación, se describirá la función de cada componente y las consideraciones del mismo.

2.1 Frameworks SSP

El componente de Frameworks SSP hace referencia a las herramientas de señales sociales que el desarrollador quiera implementar y conectar con el framework propuesto en este artículo, está contemplado que el desarrollador pueda integrar diferentes herramientas como las vistas en la sección 1 de este artículo, algunas de estas dan información de voz, otras dan información de video y análisis facial, otras información mixta entre otros tipos de información que procesan, el framework propuesto permite de manera flexible que sea posible de utilizar la herramienta deseada y de combinar sus salidas con las de otras herramientas de señales sociales, este proceso de integración y combinación se realiza en el componente de

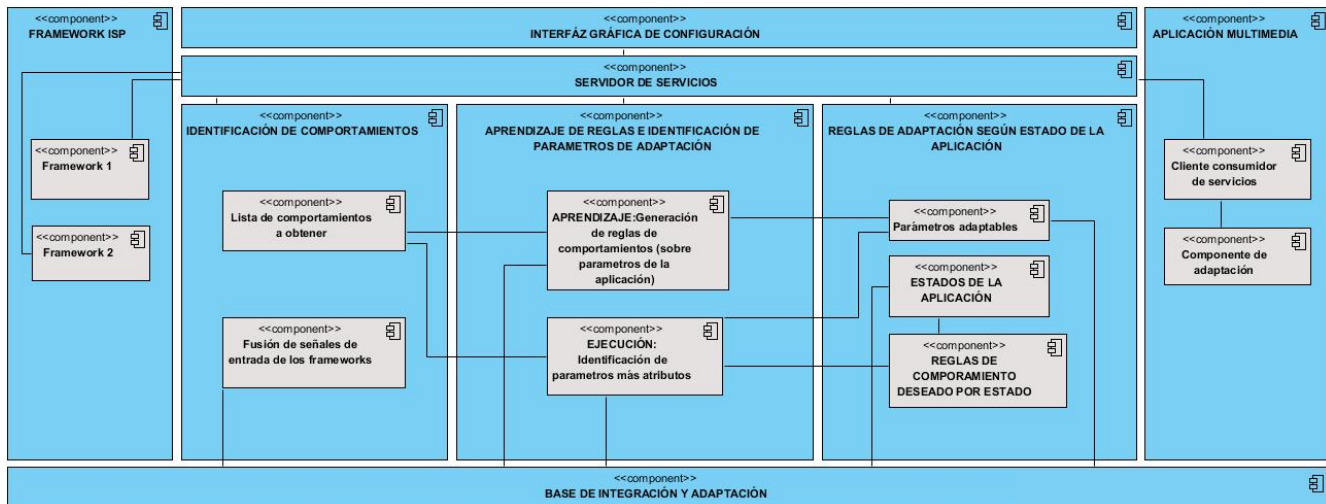


Figura 3. Diagrama de componentes de la arquitectura general de la solución propuesta.

“Identificación de comportamientos” el cual será descrito a continuación.

2.2 Identificación de comportamientos

En este componente se definen las características y la forma en que se debe realizar la comunicación con los Frameworks de señales sociales que se deseen utilizar. Dentro de este componente se encuentran 2 sub-componentes los cuales son el “*Lista de comportamientos a obtener*” y el “*Fusión de señales*”, el primer sub-componente nombrado es un servicio web el cual servirá como puerta de entrada hacia los Frameworks SSP este servicio web recibirá los valores de salida que arrojan cada uno de las herramientas SSP, este web service debe ser configurado a través de un archivo XML en el cual se especifica cada una de las posibles salidas de la herramienta y sus posibles valores. El segundo componente es el del análisis y combinación de comportamientos, con comportamientos se hace referencia a los posibles comportamientos que pueden ser concluidos de las señales sociales que se reciben por uno o varios frameworks SSP, estos comportamientos deben ser creados y definidos por el desarrollador, estos se almacenarán en el componente de “base de comportamientos que se describirá más adelante”, después de recibir las señales por medio del servicio web estas deben ser combinadas y analizadas para dar una conclusión la cual debe corresponder a uno de los comportamientos existentes en la base de comportamientos. La combinación y análisis se realiza por medio unas reglas que el desarrollador define, en las cuales se establece según las señales que ingresan y sus valores a que comportamiento de los existentes en la base de comportamientos corresponden.

2.3 Base de comportamientos

En este componente se almacenarán cada uno de los comportamientos que el desarrollador que use el framework propuesto desee según las características de su aplicación y los Frameworks SSP conectados, se crea una lista de posibles

comportamientos que deben estar relacionados a las señales que de salida que arrojan las herramientas SSP, esta lista de comportamientos será utilizada para poder concluir en un término las señales de uno o varios frameworks SSP conectados y para poder configurar según la aplicación multimedia requiera que comportamientos se quiere obtener.

2.4 Aprendizaje de reglas

Este componente es en donde se realizará un procesamiento automático, en el cual, de acuerdo a las entradas recibidas, la conclusión hecha en el componente “*Identificación de parámetros*”, el estado actual de los parámetros de la aplicación multimedia y los comportamientos esperados por la aplicación, se identifican los mejores valores que deben tener los parámetros de adaptación configurados para la aplicación multimedia. Es en este punto en donde por medio del uso del usuario se va a preñando sobre los parámetros de adaptación y su efecto sobre el usuario, para poder encontrar los parámetros adecuados para cada persona, que produzcan o se acerquen al comportamiento el cual se desee generar. Para este componente se permite que el desarrollador implemente e integre algún método de aprendizaje el cual el prefiera para identificar los parámetros según los comportamientos obtenidos y deseados. Se recomiendan técnicas como algoritmos genéticos y redes neuronales que pueden ser entrenadas rápidamente. Es criterio del desarrollador utilizar el método que se adapte más a sus necesidades y los valores que concluya.

2.5 Reglas de adaptación

Este componente se encuentra la comunicación con las aplicaciones multimedia y toda la configuración de parámetros de adaptación y sus posibles valores para adaptar la aplicación multimedia, también se encuentran las configuraciones sobre los comportamientos que se desean generar según el estado de la aplicación multimedia. La comunicación se propone que será manejada a través de un servicio web el cual entregara

información sobre los parámetros que se van ajustar y con los valores que se deben ajustar para cada parámetro y el servicio también recibirá información sobre el estado actual de la aplicación. Tiene un sub-componente en donde se describirá cada uno de los posibles parámetros de adaptación y sus posibles valores, adicionalmente en otro sub-componente se debe describir de acuerdo a el estado de la aplicación que comportamiento de los de la base de comportamiento se desea obtener, el estado de la aplicación hace referencia a la condición del nivel, tiempo, avance o caso de uso en el cual se encuentre la aplicación, este estado es ligado estrictamente a las condiciones especiales de cada aplicación.

2.6 Interfaz Gráfica

La interfaz gráfica que se propone para el framework está enfocada a dar una ayuda al desarrollador para poder configurar de una forma fácil y eficiente las diferentes variables que se encuentran en cada componente anteriormente descrito. La interfaz gráfica contará con los componentes para configurar las reglas de identificación de comportamientos, configuración de métodos de aprendizaje y clasificación, configuración de comportamientos deseados y configuración de parámetros de adaptación.

2.7 Aplicación Multimedia

Este componente es la aplicación multimedia que se desea adaptar en base a los comportamientos detectados en todo el proceso, esta aplicación tendrá una serie de parámetros configurables que van a hacer que dentro de ella cambie sus condiciones.

3. Conclusiones

De acuerdo con el framework propuesto se pretende que un desarrollador de aplicaciones multimedia sea quien decida como configurar cada uno de sus componentes de entrada para que pueda obtener un sistema que se adapte automáticamente en base a señales sociales arrojadas por el Frameworks SSP propuesto, la información obtenida por el framework concluirá en comportamientos sociales los cuales serán personalizados he irán entrenando en tiempo real las aplicaciones finales.

La propuesta permite flexibilidad en las configuraciones y en futuro ofrecerá una interfaz gráfica amigable para poder controlar las diferentes variables del framework y sus sistemas integrados. Es necesario que el desarrollador que desee usar el framework tenga conocimiento o asesoramiento para que las reglas que se definan para las señales sociales y comportamientos sean coherentes con los comportamientos sociales adecuados.

Referencias

- [1] Gelareh Mohammadi and Alessandro Vinciarelli. Towards a technology of nonverbal communication: vocal behavior in social and affective phenomena. Technical report, Idiap, 2012.
- [2] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, 27(12):1743–1759, 2009.
- [3] Alessandro Vinciarelli, Maja Pantic, Hervé Bourlard, and Alex Pentland. Social signals, their function, and automatic analysis: a survey. In *Proceedings of the 10th international conference on Multimodal interfaces*, pages 61–68. ACM, 2008.
- [4] Albert Ali Salah, Maja Pantic, and Alessandro Vinciarelli. Recent developments in social signal processing. In *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, pages 380–385. IEEE, 2011.
- [5] Johannes Wagner, Florian Lingenfelser, and Elisabeth André. The social signal interpretation framework (ssi) for real time signal processing and recognition. In *INTERSPEECH*, pages 3245–3248, 2011.
- [6] Christine Laetitia Lisetti, Cynthia LeRouge, and Fatma Nasoz. The maui project: Building multimodal affective user interfaces for everyone. In *Proceedings of the AAAI-02 Workshop "Automation as Caregiver*, pages 67–74, 2002.
- [7] Johannes Wagner, Elisabeth André, and Frank Jung. Smart sensor integration: A framework for multimodal emotion recognition in real-time. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–8. IEEE, 2009.
- [8] Stefan Scherer, Michael Glodek, Georg Layher, Martin Schels, Miriam Schmidt, Tobias Brosch, Stephan Tschechne, Friedhelm Schwenker, Heiko Neumann, and Günther Palm. A generic framework for the inference of user states in human computer interaction. *Journal on Multimodal User Interfaces*, 6(3-4):117–141, 2012.
- [9] Johannes Wagner, Florian Lingenfelser, Tobias Baur, Ionut Damian, Felix Kistler, and Elisabeth André. The social signal interpretation (ssi) framework: multimodal signal processing and recognition in real-time. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 831–834. ACM, 2013.
- [10] G Caridakis, J Wagner, A Raouzaoui, Z Curto, E André, and K Karpouzis. A multimodal corpus for gesture expressivity analysis. *Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality 18 May 2010*, page 80, 2010.
- [11] Stephen W Gilroy, Marc Cavazza, Rémi Chaignon, Satu-Marja Mäkelä, Markus Niranen, Elisabeth André, Thuriid Vogt, Jérôme Urbain, Hartmut Seichter, Mark Billinghurst, et al. An affective model of user experience for interactive art. In *proceedings of the 2008 international conference on advances in computer entertainment technology*, pages 107–110. ACM, 2008.
- [12] Stefan Scherer, Giota Stratou, Marwa Mahmoud, Jill Boberg, Jonathan Gratch, Albert Rizzo, and Louis-Philippe

Morency. Automatic behavior descriptors for psychological disorder analysis. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–8. IEEE, 2013.

- [13] Phil Chou, Francesco GB De Natale, Enrico Magli, and Eckehard Steinbach. Trends in multimedia signal processing. *IEEE Signal Processing Magazine*, 28(6):200, 2011.