

**SISTEMA DE TRADUCCIÓN SIMULTÁNEA DE LENGUAJE DE SEÑAS A VOZ  
MEDIANTE UNA INTERFAZ NATURAL DE USUARIO PARA PERSONAS CON  
DISCAPACIDAD**

SISTEMA DE TRADUCCIÓN SIMULTÁNEA DE LENGUAJE DE SEÑAS A VOZ

AUTORES: Verónica Elizabeth Rodríguez Arboleda<sup>1</sup>Juan Sebastián Grijalva Lima<sup>2</sup>Yamirlis Gallar Pérez<sup>3</sup>

DIRECCIÓN PARA CORRESPONDENCIA: Universidad Internacional SEK del Ecuador, Campus Miguel de Cervantes, Área de Sistemas Informáticos. Calle Alberto Einstein S/N y 5ta Transversal, Carcelén. Quito-Ecuador. E-mail: [ygallarperez72@gmail.com](mailto:ygallarperez72@gmail.com)

Fecha de recepción: 16 - 09 - 2015

Fecha de aceptación: 22 - 10 - 2015

**RESUMEN**

El presente artículo describe el resultado de la convergencia de tres importantes escenarios en la generación de conocimiento: la utilización de tecnología, la investigación y la vinculación con la comunidad. En el ámbito tecnológico se generó un software en lenguaje C#, que mediante una Interfaz Natural de usuario NUI reconoce coordenadas corporales y detecta movimientos; un algoritmo permite la asociación entre los lexemas capturados a través de la cámara y los códigos nemotécnicos del diccionario que se encuentran almacenados en la base de datos para personas con discapacidad del habla. Con este principio, el prototipo desarrollado permite traducir de manera simultánea las señas a voz, lo cual facilita la comunicación e interacción con personas que desconocen el lenguaje gesto – espacial. Desde la perspectiva de la vinculación con la comunidad, este proyecto materializa políticas de inclusión y el deseo de atender a los discapacitados, un sector vulnerable de la sociedad, históricamente marginado y que por sus condiciones presentan enormes dificultades para acceder a servicios públicos.

PALABRAS CLAVE: NUI; Discapacidad; Traductor; Simultaneo; Señas.

**SIMULTANEOUS TRANSLATOR SYSTEM OF SIGN LANGUAGE TO VOICE  
THROUGH A NATURAL USER INTERFACE FOR PEOPLE WITH  
DISABILITIES**

---

<sup>1</sup> Ingeniera de Sistemas en Computación e Informática. Magister en Administración de Negocios. Directora-Coordinadora del Área de Sistemas Informáticos Universidad Internacional SEK.

<sup>2</sup> Ingeniero de Sistemas Informáticos y Networking, Magister en Evaluación y Auditoría de Sistemas Tecnológicos. Director de Maestría en Tecnologías de la Información.

<sup>3</sup> Doctora en Ciencias Pedagógicas. Especialista del departamento de investigación en la Universidad Internacional SEK.

## ABSTRACT

This article describes the result of the convergence of three amounts scenarios in the generation of knowledge: the use of technology, research and community outreach. In technology software generated in C # language, which through natural user interface NUI recognized body movements and detects coordinates; an algorithm allows the association between lexemes captured by the camera and dictionary mnemonic codes for people with speech that are stored in the database disability. With this principle, the prototype developed for translating simultaneously to voice the address, which facilitates communication and interaction with people who know the language gesture - space. From the perspective of community outreach, project materializes inclusion policies and the desire to care for the disabled, a historically marginalized vulnerable sector of society, and that their conditions have huge difficulties in accessing public services.

**KEYWORDS:** NUI; Disabilities; Translator; Simultaneous; Signs.

## INTRODUCCIÓN

La presencia de la tecnología en la sociedad del conocimiento ha transformado y diversificado la forma en que se comunica el ser humano; en ese contexto se han desarrollado múltiples dispositivos y estilos de interacción que permiten el intercambio de información persona-computador.

El medio directo de comunicación entre el hombre y el ordenador es la “Interfaz de usuario”, la cual ha evolucionado a través del tiempo para facilitar la interacción y crear una experiencia de uso del computador cada vez más natural.

De la Interfaz de Línea de Comandos (Command-Line Interface, CLI) cuya interacción era a través de órdenes escritas en una Consola de comandos, se pasó a la Interfaz Gráfica de Usuario (Graphic User Interface, GUI) en donde la comunicación con el ordenador se realiza en un entorno visual sencillo con imágenes y objetos gráficos que representan la información y acciones a ejecutar.

No obstante a los resultados que se evidencian en el mundo tecnológico, las personas con discapacidad en el habla enfrentan mucha dificultad de interactuar con el resto de la sociedad que desconoce el lenguaje de gestos, lo que limita sus actividades cotidianas, en especial las que requieren una fluidez de comunicación.

Gracias al desarrollo de nuevas tecnologías, hoy la interacción hombre-máquina se puede realizar con la Interfaz Natural de Usuario (Natural User Interface, NUI) que permite una interacción más intuitiva y directa a través de comandos utilizados en comportamientos humanos habituales como movimientos gestuales o lenguaje corporal.

Para el desarrollo del SISTEMA DE TRADUCCIÓN SIMULTÁNEA DE LENGUAJE DE SEÑAS A VOZ, se utilizó la Interfaz Natural de Usuario NUI.

## DESARROLLO

Si bien diversas organizaciones, dependencias y asociaciones interesadas en la calidad de vida de las personas con limitaciones sensoriales, físicas o mentales, trabajan por alcanzar cambios verdaderamente significativos en los imaginarios sociales, es necesario reconocer el eminente papel de la ciencia y de la tecnología en los procesos de transformación de esas concepciones, con lo cual la pedagogía se ha visto animada a proponer formas de intervención más focalizadas en cada sujeto, en su condición particular, acordes con sus necesidades e intereses. Esto ha significado un cambio de paradigma, por cuanto, de la visión asistencial, mecánica y rutinaria, producto de la mirada social que se tenga acerca de la condición de discapacitado.

Se avanza hacia una concepción de hombre que se le reconoce como un ser de potencialidades y posibilidades al que debe brindársele la oportunidad para desarrollar sus habilidades cognitivas a través del ofrecimiento de ambientes en los que pueda darse a la tarea de explorar activamente el medio, de investigar, de descubrir, de participar y de interactuar de manera permanente con el mundo.

En el marco desde una posición optimista, la tecnología informática y de comunicación, cumple un papel invaluable, puesto que ella puede proveerle al sujeto los soportes físicos necesarios para el desarrollo de sus potencialidades comunicativas, cognitivas y socioafectivas; a través de ella puede lograrse una aproximación a tres grandes objetivos: autonomía, independencia e inclusión, en cuanto facilita la movilidad y la intercomunicación con el mundo al sujeto con limitaciones, permitiéndole interactuar con el ordenador desde un entorno visual sencillo con imágenes y objetos gráficos que representan la información y acciones a ejecutar

Para comprender el desarrollo del Sistema de Traducción Simultánea de Lenguaje de Señas a Voz mediante NUI, se explicarán inicialmente los recursos tecnológicos utilizados:

### *Interfaz Natural de Usuario*

La interfaz natural de usuario o NUI, permite interactuar con sistemas o aplicaciones a través de señas o movimientos reemplazando otros dispositivos de entrada de uso habitual como son: teclado, ratón, lápiz óptico, joystick, entre otros.

Entre los diferentes tipos de NUI existentes se puede encontrar los siguientes (Moxo, B. A. S., 2015):

- Reconocimiento de voz.
- Reconocimiento de escritura.
- Reconocimiento visual.

- Reconocimiento de movimientos.

### *Kinect*

Kinect fue creado con el objetivo de revolucionar la experiencia que tenía el usuario al momento de controlar su consola de juegos solo con gestos corporales o comandos de voz. Este dispositivo incorpora una arquitectura de varios elementos para su correcto funcionamiento (Zhang, Z., 2012).

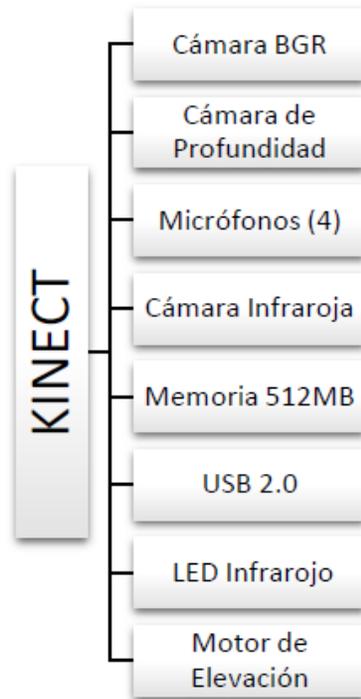


Figura 1. Arquitectura del Kinect

Las especificaciones técnicas de hardware de Kinect se detallan en la siguiente tabla:

Tabla 1. Especificaciones Técnicas del Kinect

COMPONENTES	CARACTERÍSTICAS
<b>Sensores</b>	<ul style="list-style-type: none"> <li>✓ Lentes con sensores de color y profundidad</li> <li>✓ Array de micrófonos para voz</li> <li>✓ Tilt motorizado para ajuste del sensor</li> </ul>
<b>Campo de visión</b>	<ul style="list-style-type: none"> <li>✓ Campo de visión horizontal: 57 grados</li> <li>✓ Campo de visión vertical: 43 grados</li> <li>✓ Rango de tilt físico: <math>\pm 27</math> grados</li> <li>✓ Rango del sensor de profundidad: 1.2m - 3.5m</li> </ul>
<b>Despliegue de datos</b>	<ul style="list-style-type: none"> <li>✓ 320x240 profundidad de 16-bit a 30FPS</li> <li>✓ 640x480 color de 32-bit a 30FPS</li> <li>✓ 16-bit audio @ 16 kHz</li> </ul>



Figura 2. Kinect

### *Kinect SDK (Kit de Desarrollo de Software)*

Un kit de desarrollo de software o SDK es un grupo de herramientas que permiten la creación de aplicaciones para un sistema específico, se trabajó con el SDK desarrollado por Microsoft específicamente para Kinect, la versión 1.8.

### *Lenguaje de Programación*

El SDK de Microsoft permite elegir un lenguaje de programación de la Suite de Visual Studio a partir de su versión 2012, para el sistema se seleccionó C# que es un lenguaje de programación orientado a objetos desarrollado y estandarizado por Microsoft como parte de su plataforma .NET. Su sintaxis se deriva de otro lenguaje de programación llamado C+/C++ pero con muchas mejoras, tales como: seguridad, recolección de datos, control de eventos entre otros (Hejlsberg, A., Wiltamuth, S., & Golde, P., 2006).

### *Lenguaje de Señas*

La lengua de señas, o de signos se definiría como una expresión natural de gestos que realiza el ser humano para comunicarse debido a un impedimento o discapacidad para hablar. Para el prototipo se adoptó el lenguaje homologado internacionalmente y utilizado en el Ecuador.

### *Arquitectura del Sistema*

El sistema es una aplicación de dos capas en un mismo computador, la primera es la aplicación como tal, que tiene la particularidad de utilizar la interfaz de voz propia del sistema operativo o en su defecto una genérica, lo que abre la posibilidad de tener una compatibilidad con diferentes idiomas de

salida; la segunda capa es un archivo XML o en su defecto un sistema gestor de base de datos.

### *El Equipo del Cliente*

El equipo sobre el cuál se ejecute el programa cliente tiene características muy básicas tales como:

- Procesador de 32 bits (x86) o 64 bits (x64) a 1 gigahercio (GHz) o más.
- Memoria RAM de 1 gigabyte (GB) (32 bits) o memoria RAM de 2 GB (64 bits).
- Espacio disponible en disco rígido de 16 GB (32 bits) o 20 GB (64 bits).
- Dispositivo gráfico DirectX 9 con controlador WDDM 1.0 o superior.
- Microsoft Windows 7 o superior

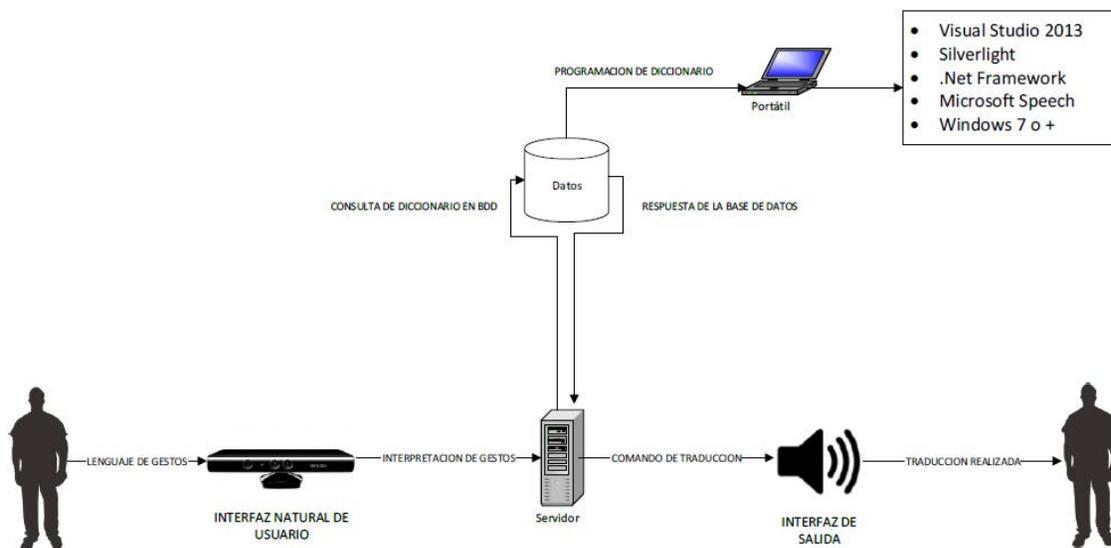


Figura 3. Arquitectura del Sistema

### *Componentes de Hardware*

El equipo cuenta básicamente con los mismos componentes que un computador de escritorio a excepción del teclado y el mouse:

- Mother Board
- Fuente de Poder
- Procesador
- Memoria RAM
- Disco duro
- Monitor
- Juego de Parlantes
- La estructura
- El Cableado

- Además como pieza clave un Kinect

### *Principios, Librerías y Algoritmos*

#### *Orientación Conjunta*

A partir del SDK 1.5, Kinect para Windows proporciona información de orientación conjunta de los esqueletos rastreados dispositivo. La orientación del hueso se proporciona en dos formas:

- Una rotación jerárquica basada en una relación de hueso definida en la estructura de la articulación esqueleto.
- Una orientación absoluta en Kinect mediante coordenadas de la cámara.



Figura 4. Componentes Internos

La información de orientación se proporciona en forma de cuaterniones y matrices de rotación para su uso en diferentes escenarios de animación.

El código en C# es:

```
private void DrawSkeletonsWithOrientations()
{
    foreach (Skeleton skeleton in this.skeletonData)
    {
        if (skeleton.TrackingState == SkeletonTrackingState.Tracked)
        {
            foreach (BoneOrientation orientation in skeleton.BoneOrientations)
            {
                // Display bone with Rotation using quaternion
                DrawBonewithRotation(orientation.StartJoint, orientation.EndJoint,
                orientation.AbsoluteRotation.Quaternion);
            }
        }
    }
}
```

```
// Display hierarchical rotation using matrix
DrawHierarchicalRotation(orientation.StartJoint,
orientation.HierarchicalRotation.Matrix)
}
}
}
}
```

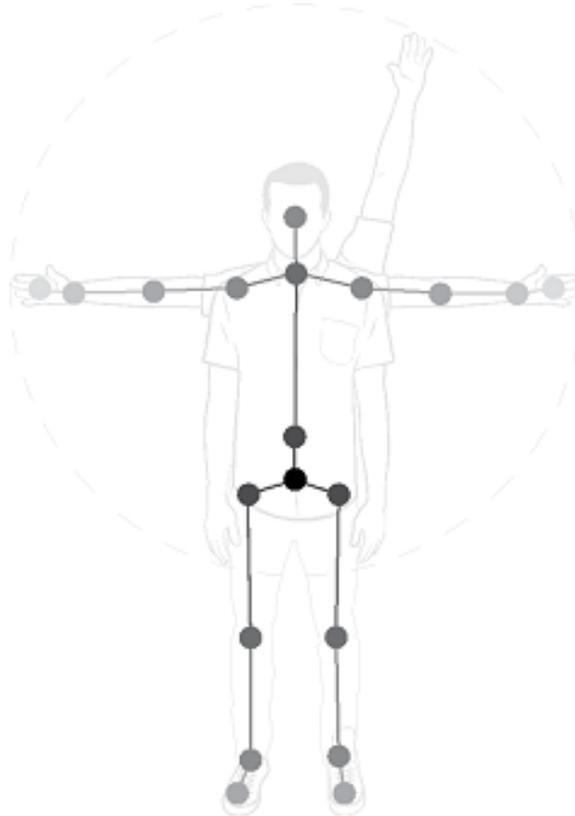


Figura 5. Jerarquía Conjunta

CENTRO DE LA CADERA				
ESPINA			CADERA IZQUIERDA	CADERA DERECHA
CENTRO DE LOS HOMBROS			RODILLA IZQUIERDA	RODILLA DERECHA
HOMBRO IZQUIERDO	CABEZA	HOMBRO DERECHO	TOBILLO IZQUIERDO	TOBILLO DERECHO
CODO IZQUIERDO		CODO DERECHO	PIE IZQUIERDO	PIE DERECHO
MUÑECA IZQUIERDA		MUÑECA DERECHA		
MANO IZQUIERDA		MANO DERECHA		

Los huesos son especificados por las relaciones entre principal y secundario según la categoría explicada. Por ejemplo, el hueso de la cadera izquierda está limitado por la articulación de la cadera central (principal) y la cadera izquierda (secundario).

En cambio la rotación jerárquica proporciona la cantidad de rotación en el espacio 3D del hueso principal para el punto secundario. Esta información indica cuánto puede girar en el espacio 3D la dirección del hueso en relación con el punto principal. Esto es equivalente a considerar la rotación del eje cartesiano de referencia en el espacio de objetos de puntos principales y de hueso para el espacio objeto secundario, teniendo en cuenta que el hueso se encuentra en el eje y de su espacio objeto.

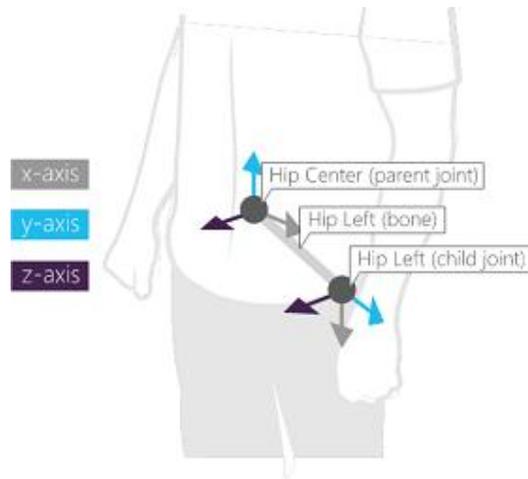


Figura 6. Jerarquía Heredada

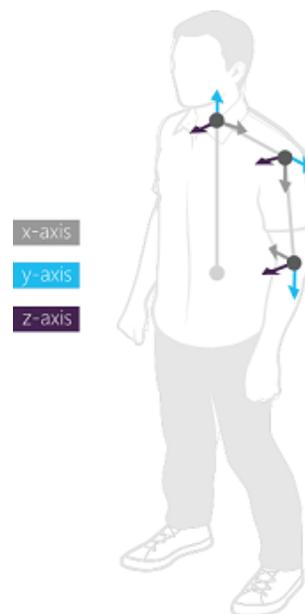


Figura 7. Rotación Heredada de Huesos

En la definición jerárquica, la rotación de la articulación de la cadera central proporciona la orientación absoluta de la persona con discapacidad en coordenadas espaciales mediante la cámara. Esto asume que el espacio entre el

usuario objeto tiene el origen en la articulación de la cadera central, el eje y es vertical, el eje x está a la izquierda, y el eje z se enfrenta a la cámara.

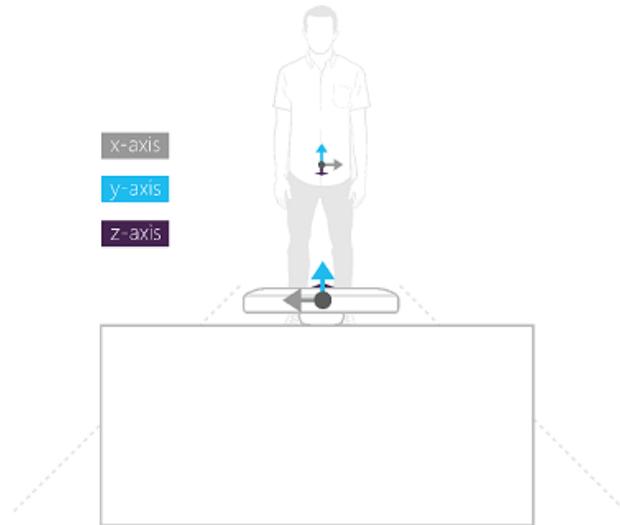


Figura 8. La Orientación Absoluta de la persona con discapacidad, inicia en la articulación de la cadera central

Para los casos de personas que se encuentran en silla de ruedas, se usará el seguimiento de un esqueleto en el modo sentado, la raíz de las articulaciones se convierte en la articulación del centro de los hombros. La rotación del resto de las articulaciones será desactivada.

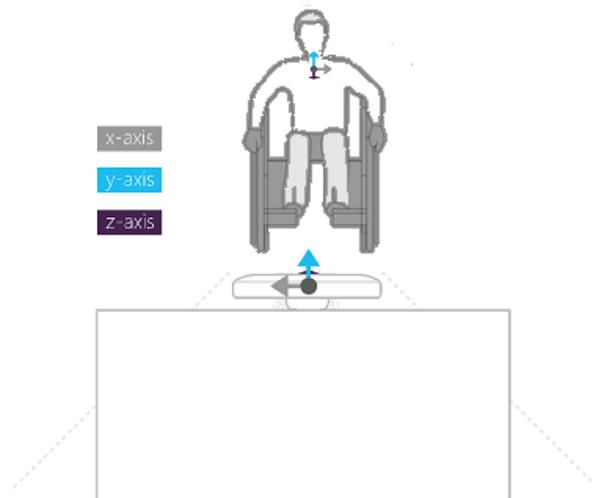


Figura 9. La Orientación Absoluta en Personas en Silla de Ruedas

La Orientación absoluta proporciona la orientación de un hueso en el espacio de la cámara 3D. La orientación de un hueso es relativa a la articulación del punto secundario y la articulación de la cadera central mantiene la orientación de la persona con discapacidad. Las mismas reglas aplican para el modo de sentado y articulaciones no rastreadas, esto en el caso de tener alguna persona que pudo haber sufrido de amputaciones.

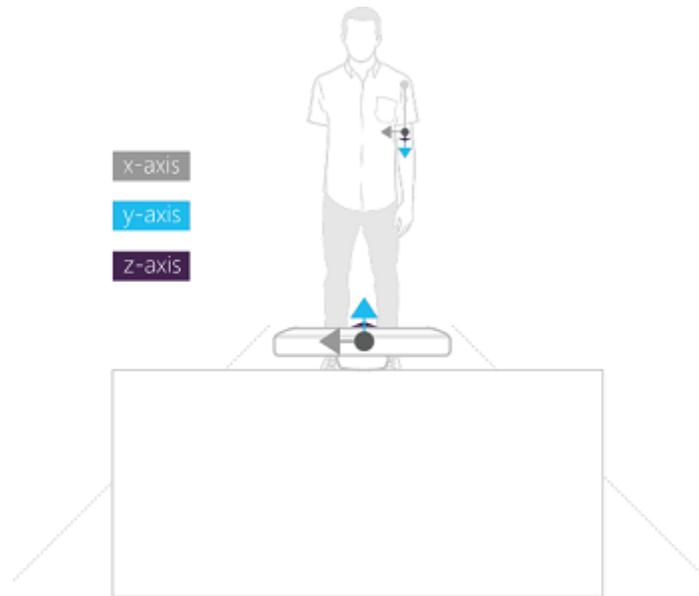


Figura 10. La Orientación Absoluta de la personas con amputaciones

### *Funcionamiento del Sistema Traductor*

El sistema inicia con el comando programado de levantar una de las manos, esta acción ejecuta el procedimiento de inicializar las variables, cargar las librerías y verificar las referencias. El sistema identifica la presencia del dispositivo (KINECT), se inicializan las librerías que permiten: reflejar en la pantalla el esqueleto o puntos de referencia del usuario, la imagen superpuesta al mismo y el sonido. Inicia el modo de lectura o identificación en el que se realiza una constante comparación de los movimientos del esqueleto con el diccionario pre configurado; al detectar una coincidencia en las coordenadas encontradas con las coordenadas almacenadas envía la orden al interprete, convirtiendo esta orden en una frase que se escucha a través de los parlantes.

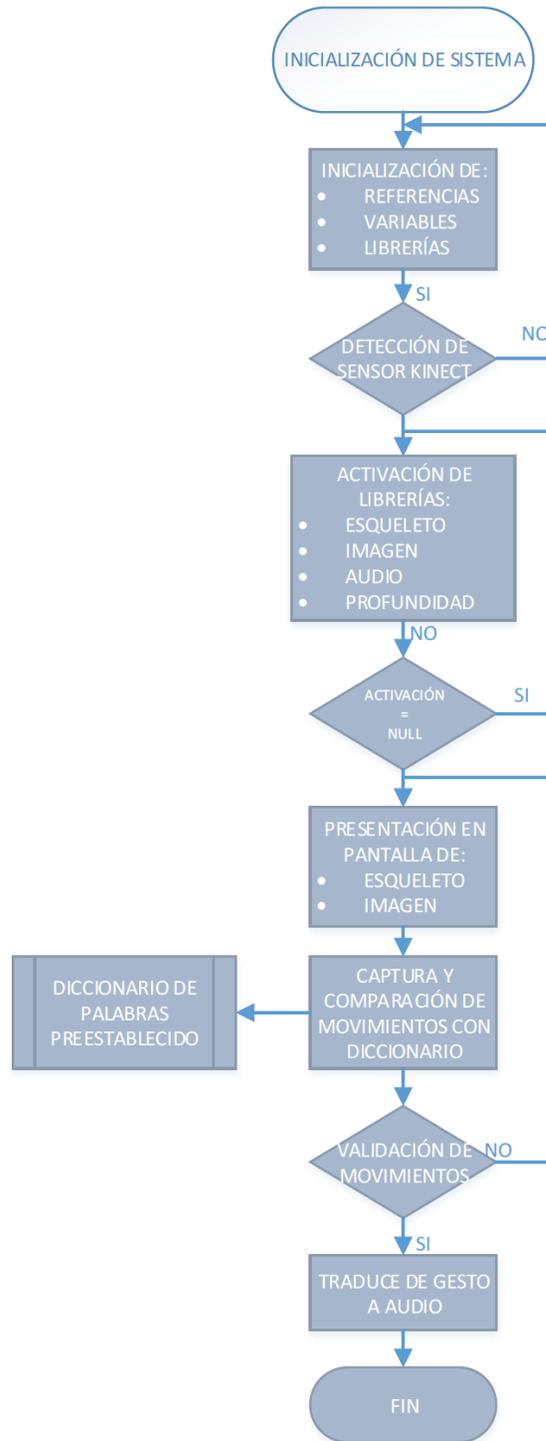


Figura 11. Proceso del Sistema Traductor

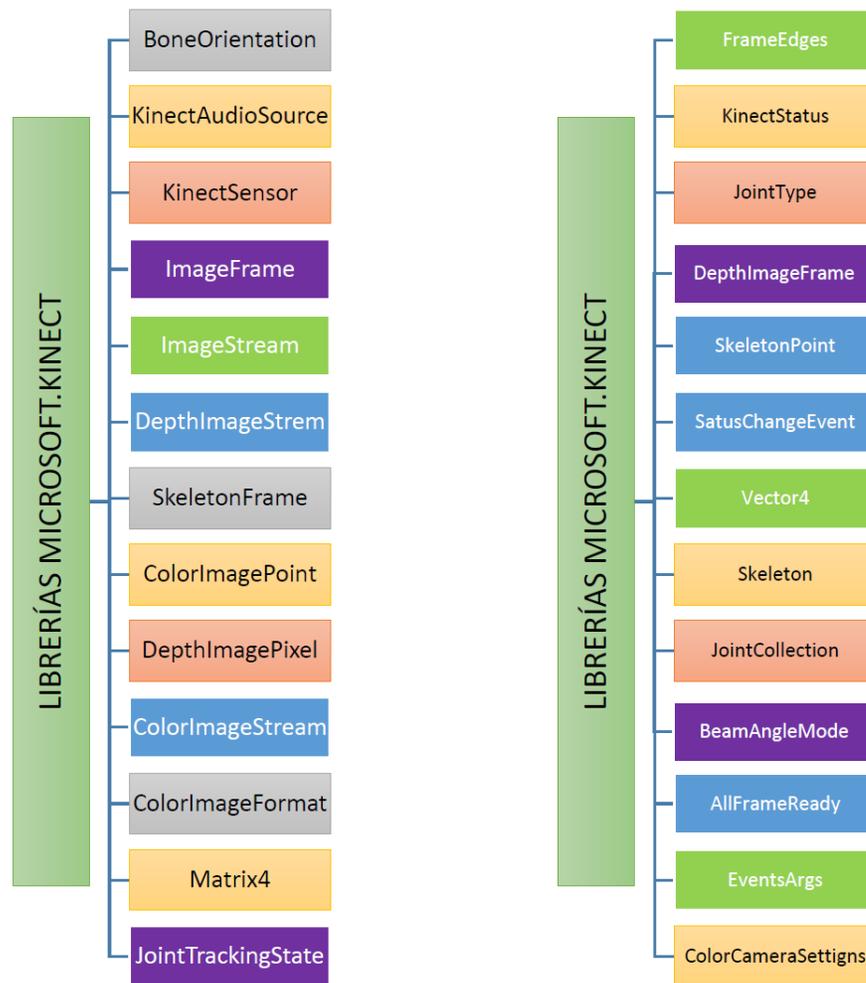


Figura 12. Librerías Base del Kinect

Estas son las librerías de dispositivo Kinect, se estima que Microsoft invirtió alrededor de 20 años de investigación para obtener la primera versión del Kinect.

### *Algoritmos*

El algoritmo de Identificación de Puntos Críticos establece cuales son los puntos que el sistema identifica mediante el análisis dimensional del cuerpo, y permite la superposición del llamado “esqueleto”, el cual será la base de la identificación de los movimientos.

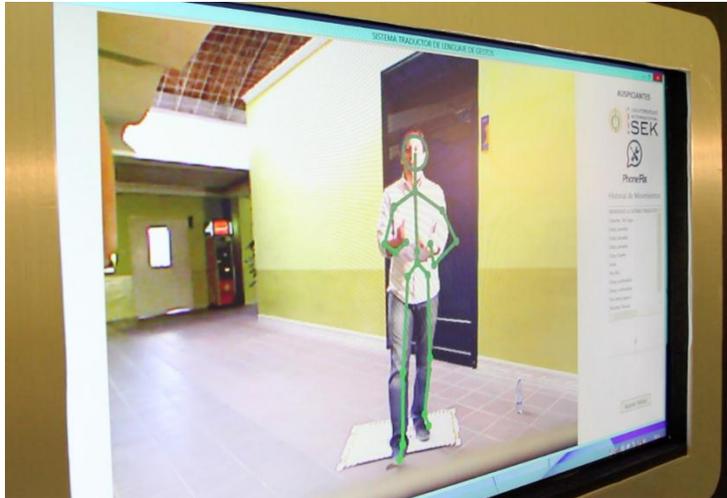


Figura 13. Ejemplo de Esqueleto identificado por el Sistema Traductor

El esqueleto puede tener dos estados, uno de "seguimiento" y otro de "única posición".

Un esqueleto en estado de seguimiento proporciona información detallada acerca de la posición, en el campo de visión, de veinte articulaciones del cuerpo de la persona con discapacidad, en relación a la cámara.

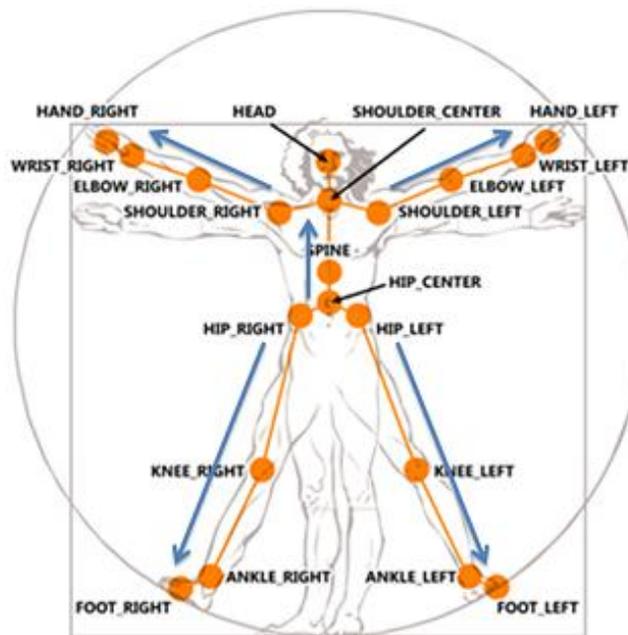


Figura 14. Mapa de Movimientos del esqueleto

Un esqueleto con un estado de "posición única" tiene información sobre la posición del usuario, pero no hay detalles sobre las articulaciones.

Como es de suponerse, también existe un modo de seguimiento para personas que se encuentren en una silla de ruedas, se ha diseñado para identificar a las

personas que están en esta posición, o cuya parte inferior del cuerpo no es del todo visible para el sensor.

El sistema desarrollado utiliza el identificador de seguimiento; el modo por defecto está orientado a personas que se encuentran paradas, ya que se puede utilizar un mayor número de combinaciones, pero se puede ajustar el sistema a las necesidades específicas de una persona de acuerdo a su discapacidad.

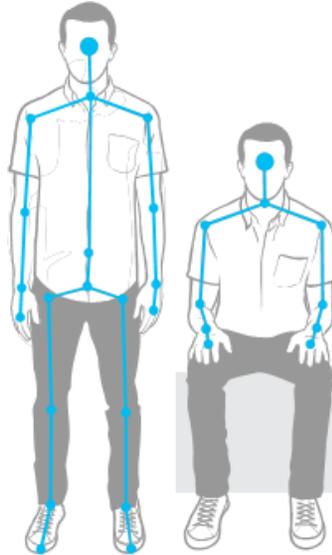


Figura 15. Puntos Programados en el Diccionario del Sistema Traductor

El modo predeterminado para personas de pie rastrea veinte articulaciones esqueléticas; en el modo sentado solo se puede dar seguimiento a diez articulaciones ubicadas en la parte superior del cuerpo (hombros, codos, muñecas, brazos y cabeza).

En el modo predeterminado o por defecto del sistema se detecta al usuario en función de la profundidad mientras que en el modo sentado, se utiliza el movimiento para detectar al usuario y distinguirlo de la silla.

Para ser reconocido en el modo por defecto basta que los usuarios se paren frente al Kinect, para el modo sentado, los usuarios deben reducir la distancia con el dispositivo e inclinarse hacia adelante o mover sus manos o el cuerpo. Si un usuario está sin moverse, es más difícil rastrearlo.

La información de seguimiento del esqueleto se puede ajustar para minimizar la trepidación y estabilizar las posiciones conjuntas con el tiempo. En el sistema se ha incluido una barra en la parte inferior de la pantalla para modificar esta característica y debe ser manejada mediante un dispositivo de entrada como un ratón o teclado.

El SDK del Kinect para Windows proporciona un mecanismo para suavizar las posiciones conjuntas, las aplicaciones pueden permitir suavizado y el sensor se ajuste la posición de las articulaciones calculados en ese marco de acuerdo al comportamiento de alisado deseado.

Respecto a la predicción, el número de fotogramas deben tener valores mayor que o igual a cero. Los valores superiores a 0,5 probablemente conducirán a sobre carga de datos cuando la persona con discapacidad se mueve rápidamente.

Se ha explicado como el sistema realiza las funciones de entrada y obtiene todos los parámetros de lectura que requiere. Pero es importante aclarar que en el sistema previamente se ha almacenado las coordenadas correspondientes a las señas del diccionario de Lengua de Señas Ecuatoriano “Gabriel Román” para que sean contrastadas con los datos de entrada o señas detectadas. Las señas pregrabadas en el diccionario están a la vez asociadas a una determinada palabra o frase que indica el movimiento realizado.

Mediante un algoritmo se compara las coordenadas o posiciones almacenadas en el diccionario con las obtenidas en la detección de posturas, si se obtiene una coincidencia se emite una alerta de salida, esta alerta va acompañada con la frase o palabra asociada a la seña.

El algoritmo de asociación de narración, utiliza el narrador del sistema operativo, para que lea el texto de notificación de la seña identificada.

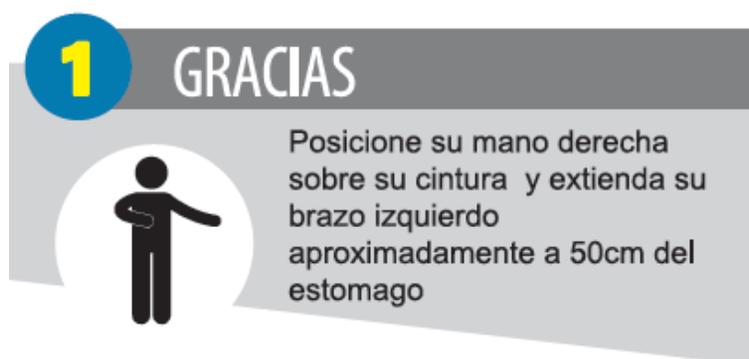


Figura 16. Ejemplo de seña grabada en el diccionario del Sistema Traductor

El siguiente código corresponde al segmento del programa que compara la seña “gracias” del ejemplo con el diccionario de señas.

```
#region 1 Gracias OK OK
int contadorgracias = 0;
if (manoizq.Position.X < codoizq.Position.X && manoizq.Position.Y > codoizq.Position.Y)
{
    contadorgracias = 1;
}
else
{
    contadorgracias = 0;
}
if (contadorgracias == 1)
{
    if (manoder.Position.X < columna.Position.X && manoder.Position.Y >
    caderacen.Position.Y && manoder.Position.Y < columna.Position.Y)
```

```

{
txtHabla.Text = ("Muchas Gracias");
BotonHablar_Click(null, null);
HistorialLista.Items.Add(txtHabla.Text);
}
contadorgracias = 0;
}
#endregion

```



Figura 17. Ejemplo de Utilización del Sistema Traductor

La idea de compartir la esencia del traductor de señales, si bien asume la tecnología como soporte esencial para el desarrollo de las personas con discapacidad, considera que ésta debe entrar a formar parte del proceso de humanización del ser con necesidades especiales, y las ayudas que de ella se deriven deben ser utilizadas como mediadoras para alcanzar una mejor calidad de vida y lograr el mayor grado de independencia posible, según la severidad del problema; ellas facilitan la accesibilidad al diálogo, a la información, a la educación, a la comprensión, es decir, a las metas y propósitos de cada hombre en interacción con la sociedad.

## CONCLUSIONES

La combinación de investigación, uso recursos tecnológicos y aplicación del conocimiento sobre Ingeniería de Software permitieron desarrollar el SISTEMA TRADCUTOR DE LENGUAJE DE SEÑAS que servirá para mejorar la comunicación con personas que por alguna razón tienen discapacidad de habla.

El modo de desarrollo preestablecido o por defecto en el Sistema está orientado a personas que se encuentran de pie, pero la facilidad de parametrización de

los movimientos, que en la programación son considerados como una concatenación de ciertas acciones con un determinado orden dentro del diccionario, permite ajustar el sistema a las necesidades específicas de la persona con discapacidad.

Para detectar los movimientos de las personas que se encuentran sentadas el sistema requiere más fotogramas por segundo y más rápido procesamiento de tramas lo que exige un mayor rendimiento y afecta a su tiempo de respuesta.

El código del sistema puede ser modificado para identificar otras señas a más del Diccionario de Lengua de Señas Ecuatoriano “Gabriel Román” utilizado en la investigación, ya que bastará con programar las coordenadas de la seña requerida y asociarla a una palabra o frase que indica lo que significa el movimiento.

Al utilizar el narrador del Sistema Operativo para leer el texto asociado a la seña identificada, se puede aprovechar la diversidad de idiomas que ofrece el sistema operativo que se utilice, con lo cual se puede programar fácilmente para que el SISTEMA TRADUCTOR DE LENGUAJE DE SEÑAS sea utilizado en cualquier país.

Sin duda alguna el proyecto de investigación constituye un producto académico de gran aporte a la educación y la sociedad.

#### BIBLIOGRAFÍA

Aznárez-Mauleón, M. (2000). Comunicación no verbal y discurso en la fraseología metalingüística con hablar o decir en español actual.

Khoshelham, K. & Elberink, S. O. (2012). Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*,

Santillán, N., & Carolina, K. (2015). *Aportes lingüísticos para la sistematización de la lengua de señas de Quito* (Doctoral dissertation, Quito/PUCE/2015).

Zhang, Z. (2012). Microsoft Kinect sensor and its effect. *MultiMedia, IEEE*

Fuentes electrónicas.

Consejo Nacional para la Igualdad de Discapacidades (CONADIS). Diccionario de Lengua de Señas Ecuatoriana Gabriel Román. Ecuador. Disponible: <http://www.consejodiscapacidades.gob.ec/diccionario-de-lengua-de-senas-ecua>

Microsoft. Tracking Users with Kinect Skeletal Tracking. Disponible en: <https://msdn.microsoft.com/en-us/library/jj131025.aspx>