Acta de Correcciones al Proyecto de Grado
Ingeniería Electrónica.

Fecha: 8 de Septiembre del 2021

Autor: Johan Sebastian Blandón Ordóñez

Nombre del Proyecto de Grado: Interfaz de voz humano-robot para controlar un brazo robótico UR3.

Directores: Msc. José Hernando Mosquera de la Cruz
Msc. Juan David Contreras Pérez

Como indica el artículo 2.27 de las Directrices de Trabajo de Grado, he verificado que los estudiantes indicados arriba han implementado todas las correcciones que los Jurados del Proyecto de Grado definieron que se efectuaran, como consta en el Acta de Calificación correspondiente.

Firma de Director(a) del Proyecto de Grado
Hernando Mosquera

Firma de Codirector(a) del Proyecto de Grado
Juan David Contreras

Pontificia Universidad JAVERIANA Cali

Facultad de Ingeniería y Ciencias

Nota de Aceptación

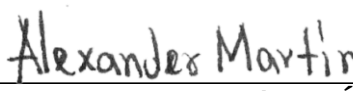
Aprobado por el Comité de Trabajo de Grado en cumplimiento de los requisitos exigidos por la Pontificia Universidad Javeriana para optar el título de Ingeniero de Sistemas y computación.

Dr. Hernán Camilo Rocha
Decano de la Facultad de Ingeniería y Ciencias

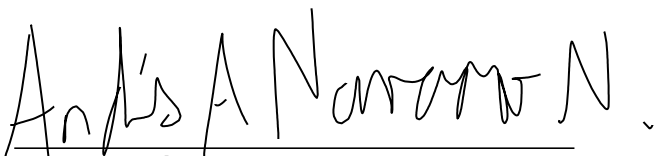
Dr. Luis Eduardo Tobón Llano
Director Carrera Ingeniería Electrónica.

Msc. José Hernando Mosquera
Director(a) Trabajo

Msc. Juan David Contreras Pérez
Codirector(a) Trabajo



Dr. Alexander Martínez Álvarez
Jurado 1



Dr. Andrés Adolfo Navarro Newball
Jurado 2

PONTIFICIA UNIVERSIDAD JAVERIANA CALI
FACULTAD DE INGENIERÍA Y CIENCIAS
DEPARTAMENTO DE ELECTRÓNICA Y
CIENCIAS DE LA COMPUTACIÓN

Interfaz de voz humano-robot para controlar un brazo robótico UR3

TRABAJO DE GRADO PARA OPTAR EL TÍTULO DE INGENIERO ELECTRÓNICO

Autor: Johan Sebastián Blandón Ordóñez

Director: M. Sc. José Hernando Mosquera de la Cruz

Codirector: M. Sc. Juan David Contreras Pérez

Santiago de Cali, 8 de septiembre de 2021
Colombia

Santiago de Cali, 8 de septiembre de 2021.

Señores
Pontificia Universidad Javeriana Cali.
Dr. Luis Eduardo Tobón Llano
Director Carrera de Ingeniería Electronica
Cali.

Cordial Saludo.

Por medio de la presente nos permitimos informarle que el estudiante de Ingeniería electrónica Johan Sebastián Blandón Ordóñez (cod: 8919690) trabaja bajo nuestra dirección en el proyecto de grado titulado "**Interfaz de voz humano-robot para controlar un brazo robótico UR3**" el cual consideramos apto para ser presentado y sometido a consideración del jurado.

Atentamente,

M. Sc. José Hernando Mosquera de la Cruz
Director de Trabajo de grado

M. Sc. Juan David Contreras Pérez
Codirector de Trabajo de grado

Santiago de Cali, 8 de septiembre de 2021.

Señores
Pontificia Universidad Javeriana Cali.
Dr. Luis Eduardo Tobón Llano
Director Carrera de Ingeniería Electronica.
Cali.

Cordial Saludo.

Tengo el placer de presentar ante usted el proyecto de grado titulado como: "**Interfaz de voz humano-robot para controlar un brazo robótico UR3**", para se sometido a consideración del jurado.

Espero que este proyecto reúna los requisitos académicos necesarios para su aprobación.

Atentamente,

Johan Sebastián Blandón Ordóñez
C.C. 1061088515 de Florencia, Cauca
Código: 8919690
blandonsebas9715@javerianacali.edu.co

Dedicatoria

Dedico este trabajo de grado a todas aquellas personas que me apoyaron en mi sueño de ser ingeniero electrónico. Familiares, amigos, compañeros, docentes, laboratoristas. En especial a mi madre que ha sido siempre mi sostén, mi padre, mis hermanos que los amo y que espero que ellos también cumplan sus sueños.

Agradecimientos

Agradezco a Dios por darme la sabiduría para culminar esta etapa. A mi madre Maria Graciela Ordoñez y mis hermanos por animarme en momentos de dificultad, a mi tía que fue una madre más en estos últimos 5 años, mi primo gracias por sus consejos y mi tío, gracias a ellos por la compañía en esta etapa tan importante de mi vida. A los laboratoristas por darme apoyo y asesorías, a los profesores por ayudarme a ser mejor persona y mejor profesional, en especial a mi director y co-director que me acompañaron durante todo el proceso.

Índice

Resumen **11**

Introducción **13**

Objetivos **15**
0.1. Objetivo General . . . 15
0.2. Objetivos Específicos . . . 15

1. Marco Teórico **16**
1.1. Procesamiento digital de señales (PDS) . . . 16
1.2. Procesamiento digital de audio (PDA) . . . 17
1.2.1. Generación de voz . . . 18
1.2.2. Reconocimiento automático del habla (Automatic Speech Recognition - ASR) 18
1.3. Sistema de diálogo hablado (Spoken Dialogue System SDS) . . . 20
1.4. Interacción humano robot (Human-Robot Interaction HRI) . . . 22
1.5. Colaboración humano robot (Human-Robot Collaboration HRC) . . . 23
1.6. Interacción humano computador (Human-Computer Interaction HCI) . . . 23
1.7. Interfaz de voz de usuario (Voice User Interface VUI) . . . 24
1.7.1. Teoría . . . 24
1.7.2. Concepto . . . 25
1.7.3. Diseño . . . 25

2. Metodología de la investigación **27**
2.1. Etapa de identificación . . . 27
2.1.1. Caracterización de las diferentes interfaces de voz implementadas en robots . 27
2.1.2. Identificación de las principales técnicas de procesamiento de voz . . . 29
2.2. Etapa de definición . . . 30
2.2.1. Identificación de los componentes del UR3 . . . 31
2.2.2. Identificación de las capacidades del robot . . . 33
2.2.3. Dispositivos, adecuaciones y restricciones de movimiento . . . 36
2.2.4. Definición del ambiente de trabajo . . . 40
2.3. Etapa de implementación . . . 44
2.3.1. Definición del diccionario de comandos . . . 45
2.3.2. Implementación del código . . . 48
2.4. Etapa de evaluación . . . 49
2.4.1. Diseño de un protocolo de pruebas . . . 50
2.4.2. Métricas de evaluación cuantitativas . . . 53
2.4.3. Métricas de evaluación cualitativas . . . 55

3. Análisis de resultados 56

Conclusiones 71

Trabajos futuros 73

Anexos 74

Bibliografía 75

Índice de figuras

1.1. Esquema del procesamiento digital de señales . . . 17
1.2. Esquema de sistema de diálogo hablado . . . 22
1.3. Diseño de interacción humano computador . . . 24
1.4. Asistente de voz de Google . . . 25
1.5. Diagrama del diseño de una interfaz de voz . . . 26

2.1. Robot UR3 de Universal Robots(UR) . . . 31
2.2. Control Box o Caja de Control . . . 32
2.3. Teaching Pendant o mando de control . . . 33
2.4. Rendimiento del UR3 . . . 34
2.5. Especificación del UR3 . . . 34
2.6. Área volumétrica del UR3 . . . 35
2.7. Movimientos del UR3 . . . 35
2.8. Pantalla del Teaching Pendant del UR3 . . . 36
2.9. Sensor de torque y fuerza, marca Robotiq . . . 37
2.10. Wrist camara, marca Robotiq . . . 37
2.11. Gripper o pinza, marca Robotiq . . . 38
2.12. Diademas Logitech G Series G935 . . . 38
2.13. Micrófono Yeti marca Blue . . . 39
2.14. Cables presentes en todo le brazo robotico . . . 40
2.15. Plano de frente al Robot, se aprecia la caja de control o control box justo debajo de la mesa y esta la UPS de lado izquierdo de la mesa . . . 41
2.16. Plano desde arriba de la mesa del robot, se logra dimensionar la elevación del robot gracias a la plataforma movible en la que se encuentra posicionado . . . 42
2.17. Cubo Amarillo . . . 42
2.18. Cubo Blanco . . . 42
2.19. Cubo Morado . . . 42
2.20. Posición inicial . . . 43
2.21. Área de trabajo . . . 44
2.22. Simulación de espacio de trabajo en Robo Dk . . . 44
2.23. Definición de HOST y PORT . . . 49
2.24. El robot levanta el cubo (*Pick*)... . . . 51
2.25. El robot pone el cubo (*Place*) . . . 51
2.26. Ilustración prueba 1 . . . 52
2.27. Ilustración prueba 2 . . . 52
2.28. Posición inicial de las fichas para la prueba 3 . . . 53
2.29. Posición final de las fichas para la prueba 3 . . . 53

3.1. Resumen de resultados de la variable cualitativa precisión en el pick en la distancia al punto de referencia de los 8 usuarios usando el Teaching Pendant 62
3.2. Resumen de resultados de la variable cualitativa precisión en el pick en la distancia al punto de referencia de los 8 usuarios usando la interfaz de voz 62
3.3. Resumen de resultados de la variable cualitativa precisión en el pick en el agarre correcto del objeto de los 8 usuarios usando el Teaching Pendant 63
3.4. Resumen de resultados de la variable cualitativa precisión en el pick en el agarre correcto del objeto de los 8 usuarios usando la interfaz de voz 63
3.5. Resumen de resultados de la variable cualitativa precisión en el pick en la recogida de la ficha correcta de los 8 usuarios usando el Teaching Pendant 64
3.6. Resumen de resultados de la variable cualitativa precisión en el pick en la recogida de la ficha correcta de los 8 usuarios usando la interfaz de voz 64
3.7. Resumen de resultados de la variable cualitativa precisión en el place en la distancia al punto de referencia de los 8 usuarios usando el Teaching Pendant 65
3.8. Resumen de resultados de la variable cualitativa precisión en el place en la distancia al punto de referencia de los 8 usuarios usando la interfaz de voz 65
3.9. Resumen de resultados de la variable cualitativa precisión en el place en el desagarre correcto del objeto de los 8 usuarios usando el Teaching Pendant 66
3.10. Resumen de resultados de la variable cualitativa precisión en el place en el desagarre correcto del objeto de los 8 usuarios usando la interfaz de voz 66
3.11. Resumen de resultados de la variable cualitativa precisión en el place en dejar la ficha en la posición correcta de los 8 usuarios usando el Teaching Pendant 67
3.12. Resumen de resultados de la variable cualitativa precisión en el place en dejar la ficha en la posición correcta de los 8 usuarios usando la interfaz de voz 67

Índice de cuadros

2.1. Tabla con los grados de rotación de cada articulación 40
2.2. Tabla con los comandos de movimiento de la Interfaz de Voz 46
2.3. Tabla con los comandos de dirección de la Interfaz de Voz 47
2.4. Tabla con los comandos de detalles de la Interfaz de Voz 47
2.5. Tabla con las métricas de evaluación cuantitativas 54
2.6. Tabla con las métricas de evaluación cualitativas 55

3.1. Tabla con el resumen de tiempos de los 8 usuarios usando el TP, se añade valor Mínimo, Máximo, Promedio y Desviación estándar 58
3.2. Tabla con el resumen de tiempos de los 8 usuarios usando la interfaz de voz, se añade valor Mínimo, Máximo, Promedio y Desviación estándar 59
3.3. Tabla con el resumen del número de instrucciones de los 8 usuarios usando el TP, se añade valor Mínimo, Máximo, Promedio y Desviación estándar 60
3.4. Tabla con el resumen del número de instrucciones de los 8 usuarios usando la interfaz de voz, se añade valor Mínimo, Máximo, Promedio y Desviación estándar 61
3.5. Tabla con el resumen de la elección de los usuarios al preguntarles acerca de con cual de las dos interfaces se sentía más cómodo. 68
3.6. Tabla con el resumen de la elección de los usuarios al preguntarles acerca de que cual de las dos interfaces permite una mejor interacción con el robot. 68
3.7. Tabla con el resumen de la elección de los usuarios al preguntarles acerca de con cual de las dos interfaces se tenia mayor facilidad de control para el desarrollo de la actividad. 69
3.8. Tabla con el resumen de la elección de los usuarios al preguntarles acerca de que con cual de las dos interfaces se demorarían menos realizando una actividad de mayor dificultad. 70

Resumen

El estudio del reconocimiento de voz en robots colaborativos ha demostrado ser una herramienta muy importante para la interacción entre el robot y el humano, al mejorar potencialmente el uso del robot. Por lo tanto, este trabajo de grado tiene como objetivo principal el desarrollo de una interfaz de voz con comandos que permita que un brazo robótico UR3 realice tareas básicas de agarre y manipulación, puesto que, se ve limitado por la dependencia sobre personal capacitado para la programación de tareas en estos robots. La propuesta del trabajo de grado que se presenta tiene un enfoque cuantitativo y cualitativo, enfocado principalmente en la medición del rendimiento de la interfaz de voz. Se implementó una interfaz de voz en el robot colaborativo UR3 del Centro de Automatización de Procesos(CAP) de la Pontificia Universidad Javeriana Cali(PUJC), a partir de un programa con reconocimiento del habla, generación de voz y de una tabla de comandos definida, los cuales se vincularon a movimientos específicos del robot colaborativo, para que este realice tareas de manipulación simples.

La interfaz de voz implementada en el robot se pretende que dote al UR3 de capacidades diferentes a las que naturalmente ya están definidas en él y que esto permita la inclusión de personas con perfiles diferentes a los que ya se definen para trabajar con el robot, como lo son personas que entiendan el lenguaje gráfico del robot y al ser un lenguaje gráfico deben ser personas sin dificultades visuales. También, se busca que la implementación de la interfaz de voz reduzca el tiempo para programar o manejar el robot ya que al programarlo con la interfaz gráfica(Teaching Pendant) hay que programar paso a paso una tarea completa.

Palabras Clave: Robot Colaborativo, UR3, Reconocimiento del habla, Colaboración Humano-Robot, Interacción Humano-Robot.

Abstract

The study of speech recognition in collaborative robots has proven to be a very important tool for human-robot interaction, potentially improving robot usage. Therefore, this study has as its main objective the development of a commanded voice interface that allows a UR3 robotic arm to perform basic gripping and handling tasks. Since, it is limited by the dependence on trained personnel for the programming of tasks in these robots. The proposal of the degree work presented has a quantitative approach, focused mainly on measuring the performance of the voice interface. This voice interface will be implemented in the collaborative robot UR3 of the Centro de Automatización de Procesos (CAP) of the Pontificia Universidad Javeriana Cali (PUJC). From speech recognition algorithms and a defined commands table, specific movements of the collaborative robot will be linked, so that it can perform simple manipulation tasks.

The voice interface implemented in the robot is intended to provide the UR3 with different capabilities to those that are naturally already defined in it and this allows the inclusion of people with different profiles to those already defined to work with the robot, such as people who understand the graphical language of the robot and being a graphical language should be people without visual difficulties. Also, the implementation of the voice interface is intended to reduce the time to program or operate the robot, since when programming it with the graphical interface it is necessary to plan step by step what movement to do.

Keywords: Collaborative Robot, UR3, Speech Recognition, Human-Robot Collaboration, Human-Robot Interaction

Introducción

Hoy en día en la robótica industrial se está hablando de una nueva linea de brazos roboticos o robots industriales, los denominados Cobots o robots colaborativos capaces de interactuar de forma segura con personas y otros robots. Ya que en estos se integra un sistema de detección y fuerza, como también medidas activas y pasivas las cuales contribuyen a la reducción de riesgo [1]. Por consiguiente, se evita bajo ciertas condiciones, la instalación de vallados de protección como los usados con robots industriales tradicionales. También, le abre una posibilidad a la academia de adquirir estos robots con mejores costos de instalación, para ampliar sus áreas de conocimiento, como ya lo hizo la *Pontificia Universidad Javeriana Cali*(PUJC) adquiriendo un robot colaborativo UR3 de la empresa Universal Robots(UR) en el año 2019.

Por otro lado, el habla es el medio de comunicación más natural y espontáneo entre los seres humanos. Sin embargo, aún en la comunicación con las máquinas y específicamente con los robots, el hombre ha hecho uso exclusivo del lenguaje escrito o tipos de programación gráficos. Resulta natural, por tanto, extender la capacidad de comunicación hombre-máquina al mensaje oral. Además de la naturalidad y espontaneidad ya nombradas, la comunicación oral hombre-máquina presenta importantes ventajas en gran cantidad de aplicaciones, como el diálogo interactivo o la entrada de grandes cantidades de datos en la máquina.

Una de estas ventajas es que en la comunicación oral las manos y la vista del usuario quedan liberadas, pudiendo dedicarse a una tarea simultánea a la comunicación[2]. Ello ofrece posibilidades muy interesantes cuando hablamos de desarrollos de gran complejidad en que la atención visual sea muy importante. Una de las áreas en las cuales se potenciaría a grandes rasgos esta ventaja sería en la medicina, más específicamente en operaciones quirúrgicas, algunos ejemplos de implementaciones ya realizadas con asistencia robótica son, Broncoscopia robótica para lesiones pulmonares[3], el CABG(Injerto de derivación arterial coronaria o Baipás coronario) robótico[4], la cirugía robótica transoral[5] y algunos que están en estudios para una posible implementación de asistencia robótica, esofagectomía, baipás gástrico, resecciones pancreáticas y hepáticas, resección rectal por cáncer [6]

Pensando en potenciar el uso del robot UR3 de la universidad y ampliar su enfoque de trabajo se plantea realizar una interfaz de voz para el robot, que integre etapas de reconocimiento de voz, un manejo de las asignaciones mediante comandos de audio definidos y una retroalimentación por voz e impresiones en linea de comandos. Adicionalmente, esta solución cuenta con detección de puntos críticos, puntos que limitan el movimiento del robot(singularidades en robótica) y cuenta con un área de trabajo definida. Por último, en algunos casos es más eficiente controlar el robot mediante comandos de voz.

Para lo cual, el proyecto plantea una evaluación de desempeño del sistema implementado. Se evaluó la interfaz de voz con valores cuantitativos y cualitativos, valores de cantidad como lo son el

tiempo , el número de intentos y el número de instrucciones para realizar una tarea. Por parte de las cualidades de la interfaz a evaluar está la precisión en dos momentos importantes de la prueba planteada, el *pick* y el *place* (que fue la tarea seleccionada para la prueba). Finalmente, se planteó realizar una encuesta comparativa de las dos interfaces para tener una perspectiva de cada usuario acerca de la comodidad, interacción y control del robot.

Objetivos

0.1. Objetivo General

Desarrollar una interfaz de voz con comandos que permita que un brazo robótico UR3 realice tareas básicas de agarre y manipulación.

0.2. Objetivos Específicos

- Identificar las principales características de las interfaces de voz implementadas en brazos roboticos.
- Definir un diccionario de comandos de voz para la interfaz en función de una tarea seleccionada.
- Implementar un algoritmo que ayude a interpretar frases de un operario a partir de técnicas de reconocimiento de voz para asociarlas a comandos de movimiento del robot.
- Evaluar la propuesta estableciendo un protocolo de pruebas que permita definir sus alcances y limitaciones.

Capítulo 1

Marco Teórico

En este capitulo se documentan todas las bases teóricas relevantes para este trabajo de grado, como lo es el procesamiento digital de señales ya que se trata el procesamiento digital de audio cuando la interfaz hace el reconocimiento del habla. También, se documenta que es un sistema de dialogo hablado el cual provee una estructura lógica del procesamiento y las etapas que debe tener una interfaz de voz para que esta se asemeje a un procesamiento natural del lenguaje como lo hace un humano cotidianamente. Luego, se documentan la interacción humano robot, la colaboración humano robot y la interacción humano computador, áreas de investigación importantes que han sido muy relevantes y estudiadas en los últimos años gracias al gran auge de las nuevas tecnologías. Por ultimo, se explica la teoría, el concepto y el diseño de una interfaz de voz y con esto poder tener claridad a la hora de pensar en la implementación de la interfaz.

1.1. Procesamiento digital de señales (PDS)

La tecnología de procesamiento digital de señales y sus avances han impactado drásticamente nuestra sociedad moderna en todas partes simplificando cada vez más las tareas que desarrollamos los humanos. Ahora, es tan simple reproducir canciones, tomar una foto, hacer un vídeo, hacer una llamada, cosas que hace 10 años eran una locura. El PDS está en todos lados y cada día se esparce mucho más dentro de las sociedades. Pensando en esto se piensa incluir mucho más PDS en el ambiente académico de la universidad con este tipo de implementación que se busca con este proyecto de grado.

El concepto básico de PDS se ilustra en el diagrama de bloques simplificado de la Figura 1.1, que consta de un filtro analógico, una unidad de conversión de señales de analógico a digital (ADC), un procesador de señal digital (DS), una conversión de señales de digital a analógico (DAC) y un filtro de reconstrucción (anti-imagen). Este procesamiento es que se le practica a cualquier señal de que procesará de forma digital, un voltage, un audio, una imagen, un video, etc.

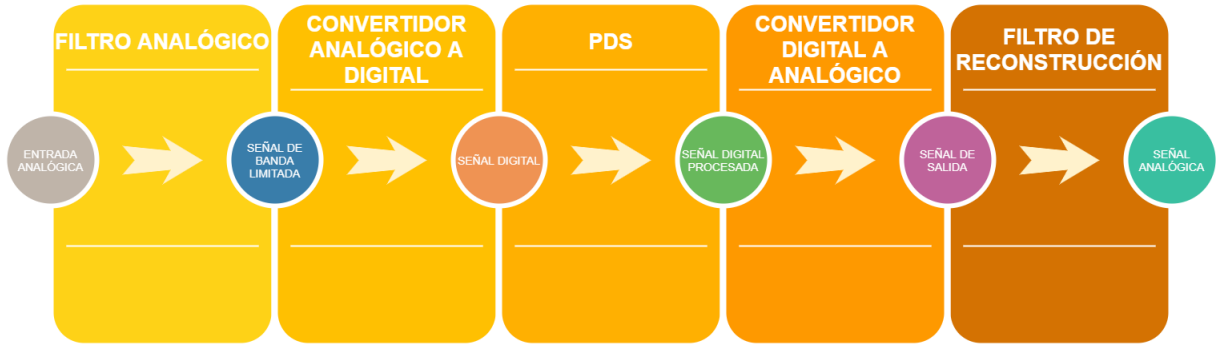


Figura 1.1: Esquema del procesamiento digital de señales

Como se muestra en el diagrama, la señal de entrada es analógica, que es continua en tiempo y amplitud, generalmente se encuentra en el mundo físico. Algunos ejemplos son señales de corriente, voltaje, temperatura, presión e intensidad de luz. Por lo general, se utiliza un transductor para convertir la señal no eléctrica en la señal eléctrica analógica (voltaje). Esta señal analógica se alimenta a un filtro analógico, que se aplica para limitar el rango de frecuencia de las señales analógicas antes del proceso de muestreo, el propósito del filtrado es atenuar significativamente la distorsión de aliasing. El paso seguido en el proceso digital de la señal es convertir la señal digital ya filtrada en una señal digital para hacer el procesamiento de la señal, se hace de forma digital por la facilidad y costo mucho más bajo para analizar y procesar la señal, se pasa a procesar la señal en algún tipo de equipo sofisticado de procesamiento(PC), luego de procesar la señal se convierte esta en una señal analógica nuevamente para poder representarla como la señal original que entró para ser procesada, y como paso final la señal se pasa por un filtro de reconstrucción de la señal.

En el campo del procesamiento digital de señales existen diferentes tipos de aplicaciones las cuales son:

Procesamiento digital de vídeo

Procesamiento digital de imágenes

Procesamiento digital de datos

Procesamiento digital de audio

De los cuales el que entra en detalle para este trabajo de grado es el procesamiento digital de audio, que se presenta a profundidad en el siguiente apartado.

1.2. Procesamiento digital de audio (PDA)

Los dos sentidos humanos principales son la vista y el oído. En consecuencia, gran parte del PDS están relacionados con el procesamiento de imágenes y audio. La gente escucha tanto música como discursos o el habla. PDS ha realizado cambios revolucionarios en estas dos áreas[7]. En este

apartado se presentan 2 frentes importantes en el PDA, que se deben tener en cuenta para entender toda la implementación de la interfaz de voz que se realizó en el proyecto de grado, estos frentes son la generación de voz y el reconocimiento del habla.

1.2.1. Generación de voz

La generación y el reconocimiento del habla se utilizan para comunicarse entre humanos y máquinas. En lugar de usar las manos y los ojos, usa la boca y los oídos. Esto es muy conveniente cuando sus manos y ojos deberían estar haciendo otra cosa, como: conducir un automóvil, realizar una cirugía o (desafortunadamente) disparar sus armas al enemigo. Se utilizan dos enfoques para el habla generada por computadora: grabación digital y simulación del tracto vocal. En la grabación digital, la voz de un hablante humano se digitaliza y almacena, generalmente en forma comprimida. Durante la reproducción, los datos almacenados se descomprimen y se vuelven a convertir en una señal analógica. Una hora entera de voz grabada requiere sólo unos tres megabytes de almacenamiento, dentro de las capacidades de incluso los sistemas informáticos más pequeños. Este es el método más común de generación de voz digital que se utiliza en la actualidad.

Los simuladores del tracto vocal son más complicados y tratan de imitar los mecanismos físicos mediante los cuales los humanos crean el habla. El tracto vocal humano es una cavidad acústica con frecuencias de resonancia determinadas por el tamaño y la forma de las cámaras. El sonido se origina en el tracto vocal en una de dos formas básicas, llamadas sonidos sonoros y fricativos. Con los sonidos sonoros, la vibración de las cuerdas vocales produce pulsos de aire casi periódicos en las cavidades vocales. En comparación, los sonidos fricativos se originan a partir de la ruidosa turbulencia del aire en las constricciones estrechas, como los dientes y los labios. Los simuladores del tracto vocal funcionan generando señales digitales que se asemejan a estos dos tipos de excitación. Las características de la cámara resonante se simulan pasando la señal de excitación a través de un filtro digital con resonancias similares. Este enfoque se utilizó en una de las primeras historias de éxito de PDS, *Speech Spell* (el habla y el deletreo), una ayuda electrónica de aprendizaje para niños ampliamente vendida.

1.2.2. Reconocimiento automático del habla (Automatic Speech Recognition - ASR)

El reconocimiento automático del habla humana es inmensamente más difícil que la generación de habla. El reconocimiento del habla es un ejemplo clásico de las cosas que el cerebro humano hace bien, pero las computadoras digitales no lo hacen como quisiéramos[8]. Las computadoras digitales pueden almacenar y recuperar grandes cantidades de datos, realizar cálculos matemáticos a velocidades vertiginosas y realizar tareas repetitivas sin aburrirse o ser ineficaces. Desafortunadamente, las computadoras de hoy en día presentan defectos cuando se enfrentan a datos sensoriales sin procesar. Enseñarle a una computadora a que le envíe una factura de electricidad mensual es fácil. Enseñar a la misma computadora a comprender su voz es una tarea importante.

El PDS generalmente aborda el problema del reconocimiento del habla en dos pasos: extracción de características seguida de coincidencia de características. Cada palabra de la señal de audio entrante se aísla y luego se analiza para identificar el tipo de excitación y las frecuencias de resonancia. Luego, estos parámetros se comparan con ejemplos anteriores de palabras habladas para identificar la coincidencia más cercana. A menudo, estos sistemas se limitan a unos pocos cientos de palabras; sólo puede aceptar el habla con pausas distintas entre palabras; y debe ser reentrenado para cada orador individual. Si bien esto es adecuado para muchas aplicaciones comerciales, estas limitaciones son abrumadoras en comparación con las capacidades del oído humano. Hay mucho trabajo por hacer en esta área, con enormes recompensas financieras para aquellos que producen productos comerciales exitosos.

A menudo se habla de las diferencias entre el reconocimiento del habla y el reconocimiento de la voz, estas diferencias pueden ser arbitrarias. Pero, básicamente la función del reconocimiento de la voz es identificar y tipificar la voz del hablante, en cambio el reconocimiento del habla distingue toda la estructura oral del usuario. Por tanto, el reconocimiento de voz permite funciones de seguridad como la biometría de la voz y el reconocimiento del habla permite transcripciones automáticas y comandos precisos. Pensando es esta cuestión se habla de reconocimiento del habla ya que la interfaz no discriminará la voz de la persona, sino solo los comandos de voz asociados a la interfaz.

1.2.2.1. Problemas relacionados al reconocimiento del habla

La dificultad para automatizar la percepción y comprensión del habla radica en su complejidad, para los humanos puede pareces simple pero los computadoras deben ser programadas como si fuesen humanos para que funcionen a la perfección estos reconocedores del habla. Se sabe que ninguno de estos procesos está integrado en la máquina en forma de algoritmos. Por ello, resulta indispensable su implementación.

- *Multiinteractividad*: Tener diferentes niveles de conciencia y / o comprensión que interactúan dinámicamente entre sí y en conjunto con otros sistemas perceptivos (como la visión) y motoros (interacción entre el aparato fonador y auditivo), etc.). Cada uno de estos niveles utiliza su conocimiento del idioma para extraer la parte correspondiente de toda la información necesaria para comprender el idioma. La estratificación más comúnmente aceptada es:

  1) Nivel acústico: en este nivel se analizan las características físicas de la señal de voz (energía, frecuencia fundamental, formantes, transiciones, etc.).
  2) Nivel fonético: en el cual se extraen los objetos sonoros elementales (fonemas, ruidos simples, etc.).
  3) Nivel léxico: donde empieza la abstracción y se determinan las estructuras simbólicas primarias (palabras o morfemas).
  4) Nivel sintáctico: en este nivel se aplican reglas para analizar la sucesión de las palabras y se comprueba su adecuación a la gramática del lenguaje, lo cual impone una determinada relación entre ellas.

5) Nivel semántico-pragmático: donde se llega a la comprensión del significado del lenguaje (estructura del discurso), se eliminan las posibles interpretaciones absurdas y se comprueba la coherencia del mensaje recibido con el conocimiento previo que de la realidad se dispone, así como del contexto en que discurre el diálogo.

Estos niveles hacen que la interactividad del sistema de reconocimiento presentes algunas dificultades y que el procesamiento en la interfaz sea más lento.

- *Continuidad*: De por si, el habla al ser una señal de audio se constituye por elementos continuos, ni los fonemas, ni las silabas, ni tampoco las palabras, se relacionan con elementos discretos, lo cual representa un trabajo arduo para la separación y análisis de cada componente del habla. No existen pausa entre los elementos y además entre más contenido tenga la señal de identificar y analizar se debe tener el en cuenta la relevancia que tiene cada palabra analizada y el tipo de contexto al que hace referencia.
- *Variabilidad*: En el habla encontramos una gran variabilidad, puesto que no existe una estructura estricta al hablar o entablar una conversación natural, es imposible que un locutor o más pronuncie exactamente dos veces la misma silaba, palabra o frase. También se presentan variables entre locutores, ya que se encuentran variables circunstanciales de entonación, amplitud, cuerpo de la voz, etc., dependen en gran medida del sexo, edad y dando espacio a variaciones en la escala de frecuencias. Así como alteraciones producidas por el estado de animo del locutor y sus condiciones físicas (cansancio, catarro, etc.) que pueden afectar la velocidad del habla o por el modo de pronunciación (susurrar, cantar, gritar, etc.) que puede afectar la escucha de la fuente de entrada -el micrófono-.

Estos problemas siempre se presentan en el reconocimiento del habla, se trabaja arduamente para reducirlos y que cada vez una interfaz de voz tenga menos problemas como este, se han tenido grandes avances con los asistentes virtuales, los cuales trabajan con servicios en la nube que hacen que su desempeño mejore considerablemente en comparación con las interfaces de voz que solo trabajan con reconocimiento del habla y procesamiento dependiendo del computador en el que este implementada la interfaz.

1.3. Sistema de diálogo hablado (Spoken Dialogue System SDS)

Un sistema de lenguaje hablado comprende el reconocimiento del habla, el procesamiento del lenguaje natural y la tecnología de interfaz humana, como el entendimiento de un dialogo[9]. Su función es reconocer las palabras que la persona dice, interpretando la secuencia de palabras para conseguir un significado en términos de la aplicación y proporcionar una respuesta idónea para el usuario. Las aplicaciones potenciales de los sistemas de lenguaje hablado van desde tareas simples, como recuperar información de una base de datos existente (informes de tráfico, horarios de las aerolíneas), hasta problemas interactivos para resolver tareas que implican una planificación y un razonamiento complejos (planificación de viajes, enrutamiento de tráfico).

Ha existido un interés renovado en el procesamiento interactivo del lenguaje natural hablado, que ya ha demostrado su utilidad para mejorar las interacciones entre humanos y computadoras[10]. Para procesar el lenguaje hablado a partir de interacciones entre humanos o entre humanos y computadoras, debemos tener en cuenta el "ruido interactivo" que implica el lenguaje hablado espontáneo. Ruido interactivo se refiere a las interjecciones, pausas, repeticiones, correcciones de errores, comienzos en falso, construcciones sintácticas o semánticas poco comunes, etc. que ocurren en el lenguaje hablado espontáneo pero no en el lenguaje escrito.

Los reconocedores de voz, que utilizamos para identificar las palabras habladas para el procesamiento del lenguaje hablado, no pueden funcionar de manera óptima y, a menudo, crean un tipo adicional de ruido. Por lo tanto, pueden hacer hipótesis de palabras incorrectas y producir oraciones gramaticales. Para continuar procesando el lenguaje hablado a pesar de estos problemas, el análisis del lenguaje hablado debe ser robusto y tolerante a las fallas[11, 10].

Existen ocho áreas clave en las que se necesita investigación básica para producir sistemas de lenguaje hablado: 1) reconocimiento del habla robusto; 2) entrenamiento y adaptación automáticos; 3) habla espontánea; 4) modelos de diálogo; 5) generación de respuestas en lenguaje natural; 6) síntesis de voz y generación de voz; 7) sistemas plurilingües; y 8) sistemas multimodales interactivos[12]. Para le creación de la interfaz de voz se tendrán en cuanta las áreas 1), 3), 4), 5), 6) que son suficientes para el tipo de implementación básica que se ha determinado, las áreas 2), 7), 8) ya generarían un plus importante para la interfaz, podrían determinarse en trabajos futuros.

Por lo general, las SDS llevan a cabo cinco tareas principales: reconocimiento automático de voz (ASR), comprensión del lenguaje hablado (SLU), Gestión de diálogo (DM), Generación de lenguaje natural (NLG) y Síntesis de texto a voz (TTS). Las tareas generalmente se implementan en diferentes módulos de la arquitectura del sistema.[13, 14, 15]. En la figura 1.2 se puede ver la estructura de un sistema de lenguaje hablado.

El objetivo del **reconocimiento del habla** es adquirir la secuencia de palabras pronunciadas por un hablante [9, 16, 17]. Es una tarea muy compleja, ya que puede haber una gran cantidad de variación en la entrada que el reconocedor debe analizar, por ejemplo, en términos de la lingüística del enunciado, el contexto de interacción y el canal de transmisión. Una vez que el reconocedor de voz ha proporcionado una salida, el sistema debe comprender lo que dijo el usuario. El objetivo de la **comprensión del lenguaje** hablado es obtener la semántica de la oración reconocida. Este proceso generalmente requiere morfológico, léxico, sintáctico, semántico, discurso y conocimiento pragmático [18, 19].

El **administrador del diálogo** decide la siguiente acción del sistema, interpretando la representación semántica entrante de la entrada del usuario en el contexto del diálogo[20, 21, 22]. En Además, resuelve puntos suspensivos y anáforas, evalúa la relevancia e integridad de las solicitudes de los usuarios, identifica y se recupera de errores de reconocimiento y comprensión, recupera información de los repositorios de datos y decide sobre la siguiente respuesta del sistema. La generación

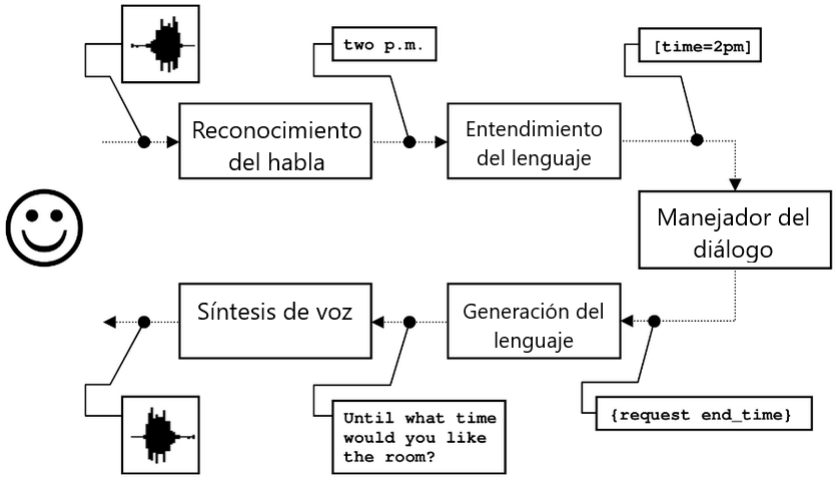


Figura 1.2: Esquema de sistema de diálogo hablado

de lenguaje natural es el proceso de obtener oraciones en lenguaje natural de la representación interna no lingüística de la información manejada por el sistema de diálogo [23, 24]. Finalmente, el módulo TTS(Text to Speech) transforma las oraciones generadas en habla sintetizada [25].

1.4. Interacción humano robot (Human-Robot Interaction HRI)

Es un amplio campo de estudio multidisciplinar con aportes en inteligencia artificial, robótica, interacción humano-computador(HCI) y más. A menudo suelen referirse a el como HRI(Human-Robot Interaction). La interacción humano robot es un campo que ha ido ganando cada vez más peso y espacio dentro del mundo de la investigación. Esto en parte se puede atribuir a la facilidad de acceso a robots de carácter doméstico u comercial que ha puesto retos en los diseños e implementación de sistemas que permitan llevar una experiencia más cómoda para las personas que hacen uso de estos. Antes de comenzar a hablar netamente de HRI hay que definir de manera correcta a qué tipo de sistemas e interfaces se está refiriendo.

Si es cierto que por medio de la inteligencia artificial se goza con mayor facilidad de experiencias donde se habla con interfaces inteligentes de voz o también conocidos, asistentes de voz cómo Cortana de Microsoft, Siri de Apple, Alexa de Amazon, Google Assistant, y M de Facebook[26]. Con estos, puedes conversar, pedir indicaciones, soluciones, preguntas, etc, dentro de ciertas limitaciones como lo es una interacción física nula, lo cual si se presenta en HRI con movimientos y en algunos casos con expresiones faciales. Estos elementos hacen que HRI se convierta en un reto de investigación que tiene como finalidad mejorar esta experiencia teniendo presentes nuevos aspectos que no se consideraban en una interacción humano computadora. Estos factores dependiendo del área en el que esté enfocado el desarrollo del robot entrarán a tener diferentes reglas o consideraciones.

1.5. Colaboración humano robot (Human-Robot Collaboration HRC)

Es un campo de investigación interdisciplinario que comprende la robótica clásica, la interacción humano-computadora, la inteligencia artificial, el diseño, las ciencias cognitivas y la psicología. Estudia los procesos de colaboración entre los humanos y los robots, ya que muchos de los trabajos actuales requieren que los humanos trabajan conjuntamente con los robots.

La colaboración humano-robot(Human-Robot Collaboration) ha ganado mucha atención en el pasar de los años. Este campo interdisciplinario que se enfoca en la colaboración de humanos y robots a medida que alcanzan objetivos compartidos [27]. Diseñado principalmente para la comunicación sobre el comportamiento humano y las respuestas del sistema, HRC abarca toda la investigación teórica y práctica asociado con el estudio, diseño y evaluación de sistemas robóticos como interactúan con los humanos [28, 29, 30].

Se ha explorado la comunicación entre robots y humanos. extensamente por muchos investigadores para una HRC óptima [31, 32, 33]. Como el primero paso hacia un HRC, los métodos de comunicación pueden idearse en términos de comunicaciones implícitas y explícitas. Específicamente, Gustavsson et al. identificó varios métodos avanzados de comunicación explícita, tales como realidad aumentada (AR) para la comunicación espacial, texto a voz (TTS) para recibir información audible de robots, reconocimiento automático de voz (ASR) para dictado de tareas y gestos para enviar comandos [32]. Además, como afirman Jan et al., La háptica La retroalimentación (táctil), la realidad virtual (VR) y el sistema de sonido audible son también enfoques factibles para cerrar las brechas de interacción entre los trabajadores y máquinas de construcción (robots), ya que la información explícita generada, como la fuerza de contacto, las señales visuales y las alertas sonoras, podría informar trabajadores sobre el comportamiento del robot durante las tareas de HRC [30].

1.6. Interacción humano computador (Human-Computer Interaction HCI)

La interacción humano-computadora consiste en el estudio, diseño, implementación y análisis de hardware y software que permite la comunicación entre usuarios y sistemas informáticos. En este sentido, una interfaz de usuario actúa como mediador entre una computadora y uno o más usuarios finales (Figura 1.3). La interfaz proporciona al usuario una representación de la sistema y con formas de interactuar con él (parte izquierda de Figura 1.3). También traduce las acciones del usuario en comandos del sistema y hace que las respuestas del sistema comprensible para el usuario (parte derecha de la Figura 1.3). El sistema o aplicación para el cual se está desarrollando una interfaz se debe distinguir de la propia interfaz ya que componen actividades y tareas diferentes, pero siempre la misma meta, la cual es funcionar en conjunto para la interacción. Esto introduce un principio básico de los sistemas informáticos interactivos: el principio de separación entre interfaz y aplicación[34].

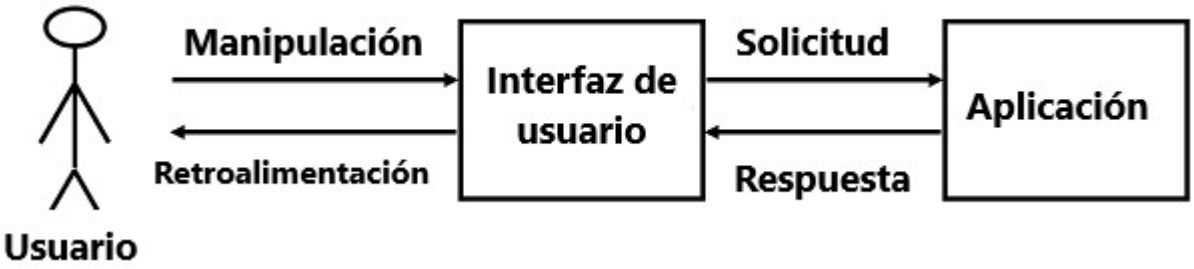


Figura 1.3: Diseño de interacción humano computador

Los seres humanos y las computadoras interactúan mediante un sistema de acceso. El sistema de acceso consta de un conjunto de componentes de hardware y software que traducir información entre el usuario y el ordenador. La traducción se produce por etapas. El usuario manipula inicialmente un dispositivo de entrada (normalmente un teclado o mouse). Estas señales pasan a través de una serie de procesos que los convierten en mensajes que la computadora puede interpretar y actuar. Las acciones de la propia computadora generan señales que luego se presentan como visuales, auditivas, o datos táctiles para que el usuario los interprete[35].

1.7. Interfaz de voz de usuario (Voice User Interface VUI)

1.7.1. Teoría

Pensando en el habla como método para comunicarse con una computadora se dimensionan ventajas importantes y su puesta en marcha tiene alcances de gran importancia para el desarrollo de la tecnología de la información. Por ende, se tienen grandes avances en los últimos años en asistentes de voz, en la actualidad su introducción al uso diario ya se puede dimensionar en los países más desarrollados. Un punto de vista de las preferencias humanas ayudan a aclarar el dominio de las interfaces visuales y manuales y las razones de elección a la interfaz de voz. Esta posición nos indica las diferencias básicas comparando dos estilos de comunicaciones. Después de todo, las circunstancias humanas también incitan a que se puede trabajar mucho en la usabilidad de los sistemas de voz. [36]

El habla como medio de comunicación con un computador o en este caso con un robot tiene muchas ventajas cuando hablamos de cooperar en las tareas humano-maquina, una de ellas puede ser un doctor y un robot en una operación, mientras que el doctor puede enfocarse en la parte de la operación mas cuidadosa, el robot puede ayudarle con las partes de la operación restante, que requiera movimientos cuidadosos o milimétricos. También, podemos hablar de una ambienta más rutinario, como una planta de producción, mientras el robot hace movimientos que demanden más peso el operario puede ayudar con las partes del proceso que demanden más cuidado y verificación. Estos dos ejemplos no demeritan el hecho de lo cuidadoso que puede llegar a ser el robot, solo que se habla de movimientos cuidadosos a los que demanden trabajo más pulido.

1.7.2. Concepto

Para que una persona se comunique con una aplicación de lenguaje hablado es necesario que esta interactué con una interfaz de voz, los elementos de esta son indicaciones de la interfaz, comandos de voz y la lógica de dialogo (también llamado flujo de llamadas). Las indicaciones, o mensajes del sistema advierten al usuario como proceder con la interacción con la interfaz, estas se muestran en forma de grabaciones o voz sintetizada que sirve como dialogo para el usuario. Los comandos de voz están definidos como las palabras a usar del usuario para las diferentes acciones que debe tener la interfaz y por ende el robot. Por tal razón, la interfaz solo podrá comprender palabra, oraciones, o frases incluidas en los comandos de voz[37]. La lógica del dialogo precisa las acciones que toma la interfaz de acuerdo a un determinado comando de voz, es decir, el usuario dijo: *"Ir a la derecha 10 centímetros"*, la lógica del dialogo seria la siguiente acción que debe hacer el robot que es moverse a la derecha del usuario 10 centímetros.

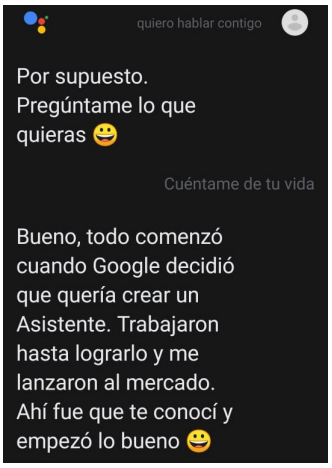


Figura 1.4: Asistente de voz de Google

En la figura 1.4, se puede ver una conversación con el asistente de voz de Google, en esto se basa una interfaz de voz con un usuario, en una máquina que puede responder como si estuviese usando un lenguaje natural, contando algo personal o mucho más general. Cabe destacar que Google usa inteligencia artificial para responder tan naturalmente a una petición y para desarrollar una interfaz así de natural gastaron mucho tiempo y dinero.

1.7.3. Diseño

Para el diseño de una interfaz de voz esta debe tener dos componentes importantes en su fundamentación, el hardware, que si se trata de una interfaz que cumpla con cualidades eficaces debe tenerse un hardware robusto, y el software que lo mismo que el hardware si la interfaz busca ser potente debe tener un software robusto. En la Fig. 1.5 se logra evidenciar el diseño de la interfaz.

Como se puede ver en la figura 1.5 acerca del diseño de una interfaz de voz, se puede ver los componentes del Hardware que son el hardware de comunicación, la comunicación de las interfaces

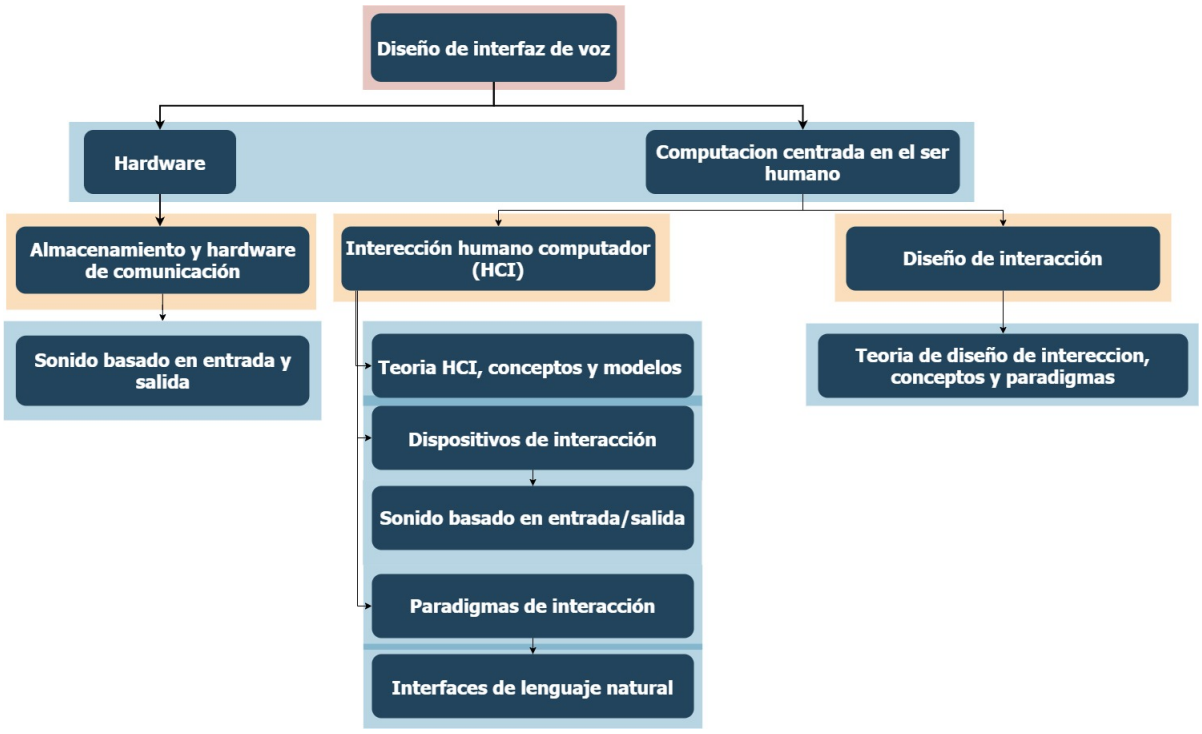


Figura 1.5: Diagrama del diseño de una interfaz de voz

y el almacenamiento, esto basado en el sonido de entrada y salida, equipos de buena adquisición de audio y reducción de ruido son indispensable, altavoces para una buena retroalimentación para tener un buen tipo de dialogo con la interfaz de habla. Por el lado del software, se habla de la interacción humano computador (HCI), la cual contiene dispositivos de interacción basados en sonidos de entrada y de salida, estos pensando en el procesamiento del audio y la generación del audio. También, en HCI se integra la interacción de paradigmas la cual propone responder a la necesidad de contar con interfaces lo más naturales posibles para el ser humano. Por ultimo, para la interacción del diseño se debe tener clara la teoría, los conceptos y los paradigmas básicos generales de las interfaces.

Capítulo 2

Metodología de la investigación

Como se tuvo previsto en el cronograma en el anteproyecto se trabajó en cuatro etapas en la metodología que estructuró el trabajo de grado. En la primera etapa se cumplió con la identificación de diferentes interfaces de voz ya implementadas en robots, se documentaron los artículos más relevantes para el estudio, también se realizó la identificación de las principales técnicas de procesamiento de voz para tener claridad de los tipos de procesamiento que se manejan en la actualidad, con esto terminó la primera etapa de la metodología y ayudó para tener claridad a la hora de abordar la siguiente etapa que fue la definición.

En la segunda etapa se identificaron los componentes del robot, se definen las capacidades del robot también los dispositivos a usar para la interfaz, las adecuaciones y limitaciones que se debieron hacer para llevar a cabo el trabajo de grado y por último en esta etapa se definió el ambiente de trabajo en el cual se especifican las condiciones ambientales del CAP, los objetos a usar, lugar de ubicación del usuario que realiza la interacción con la interfaz, etc. En la tercera etapa de la metodología, la implementación, se define el diccionario de comandos y se documenta la implementación del código, el lenguaje de programación usado, las librearía usadas y una explicación general del código de la interfaz de voz. Por último, la etapa cuatro en la cual se realizó la evaluación, se define el protocolo de pruebas, las métricas cuantitativas y cualitativas para realizar la calificación de las capacidades de la interfaz de voz.

2.1. Etapa de identificación

Para la etapa de identificación se tuvo en cuenta la literatura específicamente con trabajos relacionados con interfaces de voz con diversos robots no solo con el UR3, ya que hay pocas implementaciones documentadas (publicadas) con este robot, se hacen implementaciones de interfaces de voz en cobots en hackatones y ferias de conocimiento pero no se publican, esto hace que su referencia no tenga peso científico. En la sección 2.1.1 se pueden apreciar las diferentes interfaces de voz implementadas. En esta etapa también se documentan las principales técnicas de procesamiento de voz, en la sección 2.1.2 se aprecia esta información.

2.1.1. Caracterización de las diferentes interfaces de voz implementadas en robots

La comunicación siempre ha sido un área de investigación importante en ingeniería, en la actualidad el reconocimiento del habla es indispensable en las comunicaciones con los sistemas inteligentes

[38]. Asistentes virtuales que usan reconocimiento del habla como los antes mencionados(Alexa, M, etc), ayudan en tareas diarias, tales como programar una alarma, acceder a internet, agendar una cita, usar paquetes de información de la nube [39], como es el caso de Alexa. El reconocimiento del habla también se está llevando al campo de la robótica, algunos investigadores ya han hecho trabajos de reconocimiento del habla con robots como UR3 [1, 40], UR5 [40], RV M1 [41], robot Félix(Cuadrúpedo) [42], silla robótica [43], Doris (robot guía) [38], robot humanoide [44], Panda de FRANKA EMIKA [45] que es un robot colaborativo y Husky de CLEARPATH [45]. Estos ejemplos implican una apropiación de la implementación de interfaces de voz en los robots y actualmente más común en los cobots.

A continuación se documentan trabajos realizados con brazos roboticos especialmente y con cobots en los cuales se han implementado interfaces de voz. El primer documento relevante relacionado con el trabajo de grado es el de Samper et al. [1] en el cual se detalla el uso de ALEXA para el reconocimiento del habla, adicional a esto también usan paquetes de Amazon en la nube (Amazon WEB Service (AWS)) como, Alexa Skills Kit(ASK), Lambda y DynamoDB, estos usados principalmente para el tratamiento de la señal y la correspondencia dada a cada movimiento del robot ur3.

El trabajo de Gustavsson et al. [46], demuestra la colaboración humano robot(Human-Robot Collaboration(HRC)), combinando reconocimiento del habla y el control kinestésico. Para realizar el reconocimiento del habla usan un microfono Sennheiser ME 3 EW con una interfaz de audio Steinberg UR12 USB, también usan Microsoft Speech API 11 para el procesamiento de la señal de audio en el computador, junto con EasyModbusTCP que se usa para la comunicación cliente servidor con el UR3.

En el artículo realizado por Hofer y Strohmeier [45] trabajan con dos robots, uno colaborativo el Panda de FRANKA EMIKA y un UGV(Vehículo terrestre no tripulado) el Husky de CLEARPATH, lo interesante de este trabajo es que integran el reconocimiento del habla con un manejador en ROS(Robot Operating System), esto genera la ventaja de poder controlar los robots en múltiples lenguajes, Español, Ingles, Alemán, etc. Esto generando valor agregado al control por voz y haciendo la solución más universal.

Como podemos ver según la documentación, las interfaces de voz difieren en sus características, principalmente en los elementos que se usan en cada una. No hay una linea a seguir según la implementación ya que actualmente existen diferentes tipos de soluciones para el desarrollo de una interfaz. También, se aprecia que las interfaces de voz han sido y serán implementadas en robots, ya que estas los dotan con capacidades diferentes a las que convencionalmente estamos acostumbrados a apreciar. Estas capacidades logran que la HRC y la HRI sean áreas que predominen en las aplicaciones actuales en la industria 4.0 [47], abriendo un nuevo cumulo de posibilidades de implementación para los robots y en especial para los cobots(robots colaborativos), pueden estar presentes en áreas como la salud, academias, entretenimiento y demás, en estas podrían realizar actividades que hace 10 años ni estaban contempladas, actividades como, asistencia en operaciones, bar-robots (como un

bartender robotizado) y en academias asistiendo al aprendizaje, como lo es en este trabajo de grado.

2.1.2. Identificación de las principales técnicas de procesamiento de voz

En este apartado se documentan las técnicas más recientemente documentadas acerca del procesamiento del habla o voz, estas técnicas también son ampliamente usadas para el tratamiento de otras señales, como el análisis climático, el diagnóstico de fallos mecánicos, etc. Esta técnicas tiene similitudes ya que comparten algunas transformadas y también se relacionan mas abiertamente por ejemplo la descomposición de modo variacional es una implementación mejorada de la descomposición de modo empírica. Por ende, el uso que se le da a cada una de estas técnicas se define con la aplicación que se le de.

2.1.2.1. Descomposición de paquetes de Wavelet (Wavelet Packet Decomposition)

La descomposición en paquetes de wavelets es una técnica eficaz para analizar una señal de naturaleza no estacionaria, especialmente para la señal de voz. Es más eficiente desde el punto de vista computacional y tiene un buen rendimiento en comparación con las técnicas de procesamiento de señales en tiempo corto, como los métodos de transformada de Fourier. En el análisis de la transformación de paquetes de ondas (WPT), se aplica el tamaño de ventana variable para capturar la información de las bandas de alta y baja frecuencia capturada. Representa los resultados de paso alto y paso bajo como una generalización de la descomposición wavelet. El WPT descompone la señal en sub-bandas, lo que proporciona una buena resolución temporal y de frecuencia. Los paquetes wavelet son formas de onda indexadas por tres parámetros: posición, escala y frecuencia.

Podemos aprovechar el paquete de wavelet tomando una característica diferente como la característica estadística (media, varianza, curtosis y asimetría), la energía y la entropía del paquete de ondas. Estas características pueden utilizar en la clasificación del habla patológica, la verificación del hablante, el reconocimiento automático del habla (ASR), el reconocimiento de emociones y la clasificación del género. Recientemente, se han propuesto algunos enfoques nuevos que utilizan el análisis wavelet para analizar las señales de voz. El método de análisis wavelet se basa en la multirresolución para reflejar las interacciones no lineales entre vórtices y flujos. Se ha aplicado al ruido, la detección, la compresión, la clasificación, etc.[48].

2.1.2.2. Descomposición del modo empírico (Empirical Mode Decomposition)

La descomposición modal empírica (EMD) es una técnica de descomposición adaptativa para el tipo de señal no estacionaria. N. Huang introdujo esta técnica en 1998 [49]. Descompone la señal de voz en componentes AM-FM (modulación de amplitud y modulación de frecuencia) llamados modos o funciones de modo intrínseco (IMF). EMD es ampliamente utilizado hoy en día para descomponer recursivamente una señal en diferentes modos de bandas espectrales desconocidas pero separadas. EMD es conocido por limitaciones como la sensibilidad al ruido y el muestreo. Estas limitaciones

sólo podrían ser parcialmente abordados por intentos más matemáticos a este problema de descomposición, como la sincrosqueezing, wavelets empíricos o descomposición variacional recursiva. [50] La técnica EMD tiene una alta complejidad computacional y requiere una gran serie de datos.

2.1.2.3. Descomposición del modo variacional (Variational Mode Decomposition)

La descomposición en modo variacional (VMD) es la última herramienta de procesamiento de señales donde la señal de entrada se descompone en diferentes IMFs limitados en banda. VMD proporciona mejoras sobre WT y HHT, como ningún efecto de aliasing modal y es sensible al ruido. VMD tiene una excelente resistencia al ruido, un mejor rendimiento de descomposición y estabilidad y también se puede utilizar para la extracción de características y el diagnóstico de fallas [51].

2.1.2.4. Transformada Wavelet de Sincronización: EMD como una herramienta (Synchrosqueezing Wavelet Transform: EMD Like a Tool)

La transformada wavelet de sincronización (SST) es un método no lineal de tiempo-frecuencia basada en la transformación wavelet continua (CWT). Distribuye la energía de la señal en la frecuencia. Mitiga el efecto de la dispersión de la ondícula madre. La SST reasigna la energía en una dirección de frecuencia, lo que conserva la resolución temporal de la señal. Así, reconstruye mejor la señal. En el caso de una señal con mucho ruido, la SST ofrece una representación tiempo-frecuencia limpia. Se reduce el efecto de la mezcla de modos . El algoritmo SST se utiliza en muchas aplicaciones, como el análisis climático, el diagnóstico de fallos mecánicos, la demarcación de señales y análisis de señales de voz [50].

2.2. Etapa de definición

La familia de cobots de UNIVERSAL ROBOTS (UR) tiene cuatro opciones (UR3, UR5, UR10, UR16) diferentes de carga útil. El brazo robótico que se encuentra en la universidad es el UR3 con una carga útil de 3 Kg. Otro de los principales rasgos que tiene es que posee seis grados de libertad, una flexibilidad increíble y una fácil integración en los entornos de producción existentes, entre otras. A continuación se documenta en la sección 2.2.1 la identificación de los componentes del UR3 que viene desde fabrica al momento de comprar un UR3. También, en la sección 2.2.2 se encuentra la identificación de las capacidades del robot, en donde se documentan las especificaciones del robot, el rendimiento, la capacidad de movimiento de cada grado de libertad y demás. Además, en la sección 2.2.3 se encuentra la documentación de los dispositivos adecuaciones y limitaciones que estuvieron presentes en la implementación de la interfaz de voz para el robot UR3 de la universidad. Por último, en la sección 2.2.4 se define el ambiente de trabajo en conjunto con los elementos a usar en las pruebas.

2.2.1. Identificación de los componentes del UR3

Para la identificación de los componentes del UR3 se usó la hoja de datos del robot que Universal Robots entrega con la compra del robot o se puede encontrar en el siguiente link: https://www.universal-robots.com/media/1801288/eng_199901_ur3_tech_spec_web_a4.pdf , en el documento que se encuentra en el link se dispone de información acerca de los componentes del UR3 y de las capacidades del mismo que se documentan el la siguiente subsección.

2.2.1.1. Características UR3

El robot UR3 provee una protección IP64 la cual lo escuda contra polvo y salpicaduras de líquidos, esto ayuda a ensamblarlo en diferentes ambientes de trabajo, no solo en la academia(Como es el caso del CAP de la universidad). Adicionalmente, cuenta con una certificación ISO clase 5 para áreas limpias, es decir que el robot puede ser usado en aplicaciones como biotecnología, farmacéutica, nanotecnología y diversas aplicaciones de fabricación de tecnologías limpias[52], tiene un ruido de 70dB que es el equivalente al ruido del trafico de una autopista (Estas son las especificaciones de UR pero trabajando con el UR3 no es tan molesto el ruido que genera). Cuenta con cuatro puertos digitales dos de entrada (E) y dos de salida (S) y tiene dos entradas análogas.

En el apartado físico del robot este cuenta con 128mm de diámetro de la base el cual es necesario para saber el espacio requerido en la mesa o plataforma de trabajo para poder sujetar el robot a esta. El robot esta hecho de aluminio y polipropileno termoplástico, cuenta con un conector tipo M8 el cual sirve para conectar la herramienta de trabajo, este conector se usa específicamente con sensores industriales y cuenta con un tornillo impermeable de 3 clavijas, especial para entornos hostiles, en el caso del robot se pueden presentar movimientos bruscos y este conector evitaría que la herramienta se desconecte. Con el robot viene incluido un cable de 6 m/ 236 in el cual está alrededor de todo el brazo robótico una vez este conectada la herramienta de trabajo y con la instalación del cable el robot pesaría 11 kg/ 24.3 lbs. En la figura 2.1 se puede ver el robot de 6 grados de libertad.

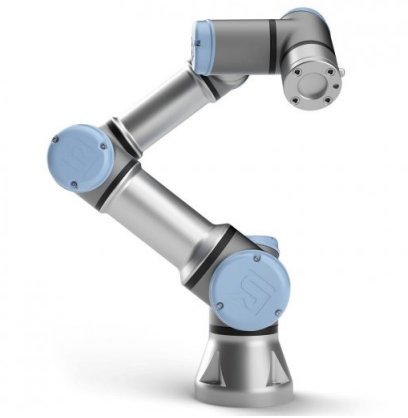


Figura 2.1: Robot UR3 de Universal Robots(UR)

2.2.1.2. CONTROL BOX

Es una caja de acero la cual es la encargada de alojar todo el control y la comunicación del robot UR3. Tiene un tamaño de 457mm x 423mm x 268mm / 18.7 x 16.7 x 10.6 in y tiene un peso de 15kg / 33.1 lbs (estas características son las de fabrica). La caja de control cuenta con un protección IP20 la cual garantiza seguridad frente a solidos con diámetro superior a 12mm y cero protección frente a líquidos. Tiene una certificación ISO clase 6 para áreas limpias, lo cual asegura una filtración de aire de penetración de alta eficiencia de 99.9 % a 0.3 micrones, tiene un ruido de trabajo menor a 65dB(10dB por encima del ruido ambiente -según OMS-[53]) lo cual equivale a un grupo de personas conversando en voz alta. Si hablamos de los puertos de entrada y de salida (I/O), tiene 16 entradas digitales, 16 salidas digitales, 2 entradas analógicas y 2 salidas analógicas, cuenta con una alimentación de 24V a 2A para estas I/O y cuenta con protocolos de comunicación TCP/IP 100Mbit, Modbus TCP, Profinet y EthernetIP. El control box funciona con una alimentación de corriente alterna de 100 V - 240 V a 50 - 60 Hz y con un rango de temperatura ambiente de 0 - 50°. En la figura 2.2 se puede ver la caja de control.

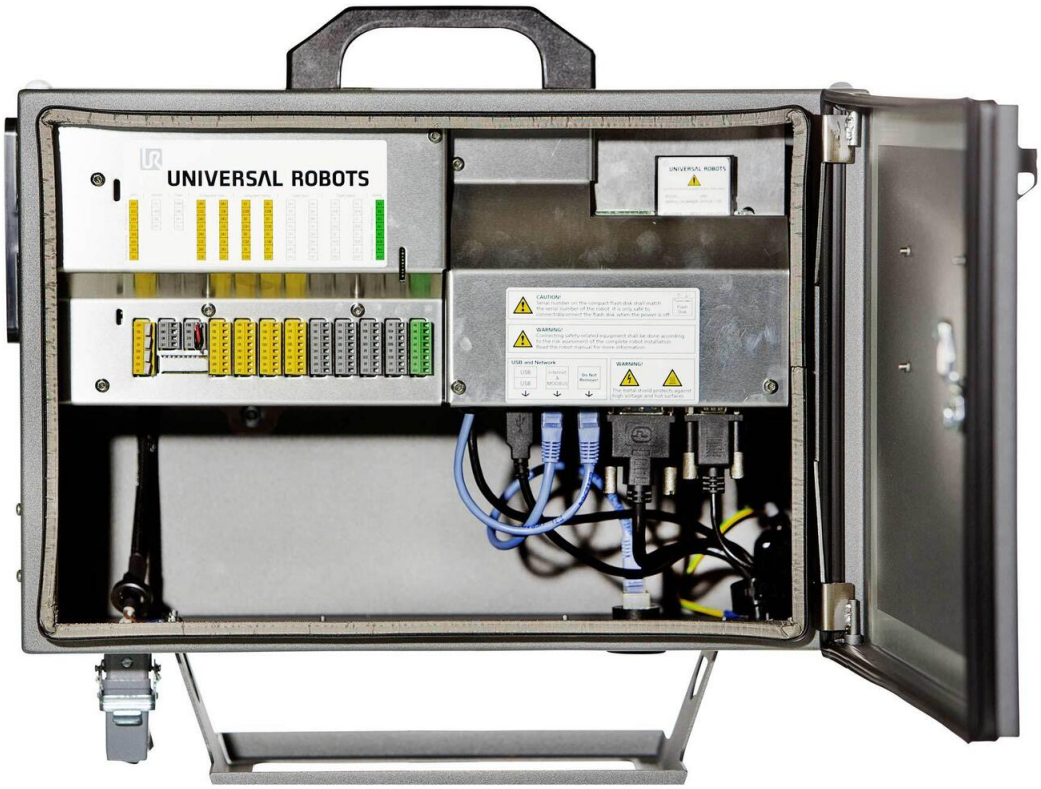


Figura 2.2: Control Box o Caja de Control

2.2.1.3. TEACHING PENDANT

Para el manejo del robot Universal Robots(UR) dispone de una tableta capacitiva llamada Teaching Pendant(TP) la cual es un tipo de interfaz humano-maquina (Human-Machine Interface HMI), el cual se refiere a un panel de control que sirve para comunicarse son una maquina, software o sistema. Cuenta con un botón de encendido del robot(Botón de Power) y un botón de emergencia(Botón Rojo) y toda su pantalla en la cual se muestra toda la interfaz gráfica para el manejo del robot. Este dispositivo viene con una protección IP20, la cual garantiza defensa frente a solidos de diámetro superior a 12mm y no provee resguardo frente a líquidos. Este dispositivo esta hecho de Aluminio y Polipropileno(PP), al tener éste materiales no tan resistentes UR provee un forro que lo protege contra golpes. El TP tiene un peso de 1.5 kg/ 3.3 lbs, lo cual no representa gran dificultad al momento de manipular el TP la primera media hora, luego ya este peso empieza a incomodar. Por ultimo, basta nombrar que cuenta con un cable de 4.5 m/ 177 in, bastante largo para poder estar alejado del robot en alguna actividad que el robot demande espacio. En la figura 2.3 se puede ver el dispositivo antes mencionado.

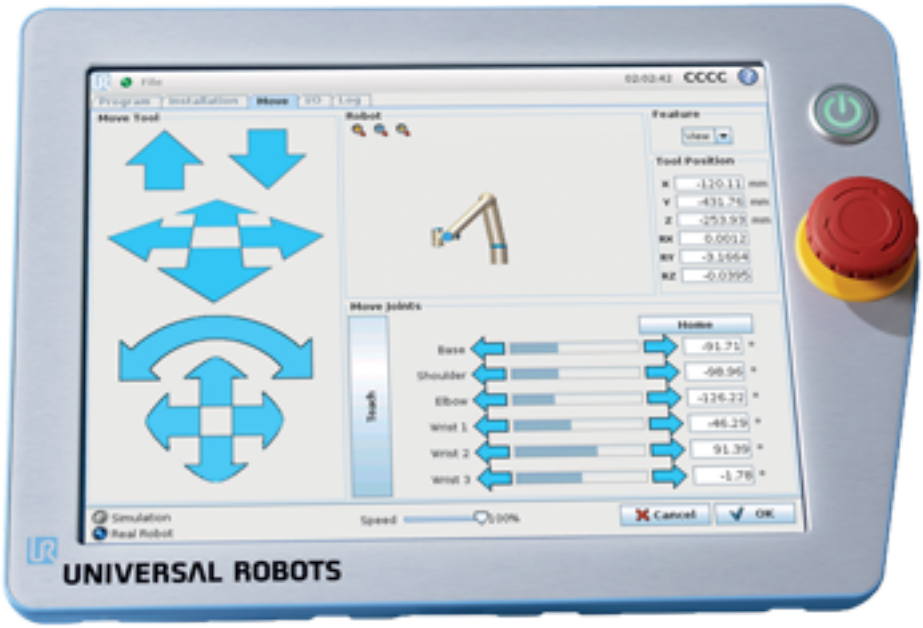


Figura 2.3: Teaching Pendant o mando de control

2.2.2. Identificación de las capacidades del robot

Resulta muy adecuado trabajar con este tipo de robots ya que, posee diferentes sensores a lo largo de su diseño que desactivará el movimiento del robot cuando se encuentre con un obstáculo. Toda su documentación técnica se puede encontrar en la página principal de la compañía, de este lugar fue donde se extrajo la mayoría de datos que se darán a continuación del UR3.

Rendimiento	
Repetibilidad	±0,1 mm / ±0,0039 in (4 mil.)
Intervalo de temperaturas	0-50°*
Consumo de energía	Mín. 90 W, estándar 125 W, máx. 250 W
Operación de colaboración	15 funciones avanzadas de seguridad regulables. Función de seguridad con certificación TÜV NORD Probado de acuerdo con las normas: EN ISO 13849:2008 PL d

Figura 2.4: Rendimiento del UR3

En la figura 2.4 se muestra las capacidades de rendimiento del UR3. Una repetibilidad de ± 0,1 mm, la repetibilidad es tomada como la capacidad del robot de regresar a un punto en específico varias veces, un valor de 0,1 mm deja ver una precisión bastante alta para esta máquina. Puede trabajar en un rango de 0 a 50° con un consumo de energía de 90 W a 250 W como medida máxima. Cuenta con certificaciones TÜV NORD y SÜD que validan las funciones de seguridad que posee el sistema del UR3.

Especificación	
Carga útil	3 kg / 6,6 lb
Alcance	500 mm / 19,7 in
Grados de libertad	6 articulaciones giratorias
Programación	Interfaz gráfica del usuario PolyScope con pantalla táctil de 12" con soporte

Figura 2.5: Especificación del UR3

En la figura 2.5 se evidencía las especificaciones sobre la carga útil máxima que puede soportar el UR3 que son 3 Kg, un alcance de 500 mm que son equivalentes a 0,5 metros. 6 articulaciones giratorias y una interfaz gráfica de 12" con soporte. Evidentemente el total de la carga útil no será usado, la carga que se maneja es por mucho menor. Por otra parte, las 6 articulaciones son excelentes para versatilidad y agilidad del robot para realizar movimientos precisos y con dificultad. Por otra parte, UNIVERSAL ROBOTS recomienda que se tengan en cuenta el área volumétrica de la base como se aprecia en la figura 2.6, que sería de 50 cm de radio para todo el volumen de trabajo, ya que de estar expuesta a vibraciones o movimientos podría llegar a provocar que el cobot trabaje de forma ineficiente.

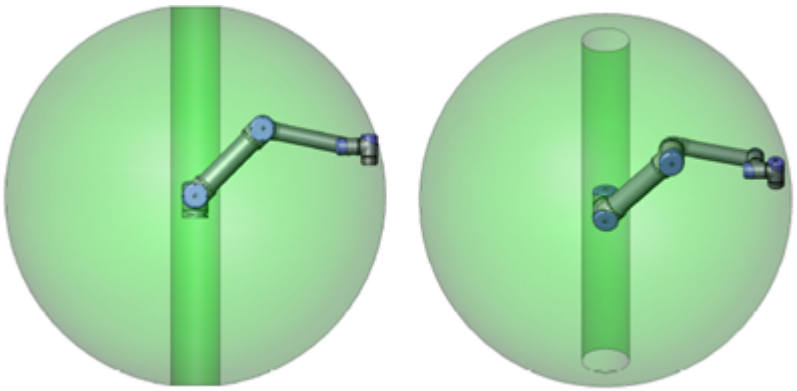


Figura 2.6: Área volumétrica del UR3

En cuanto a las capacidades técnicas del UR3 la figura 2.7 nos muestra que sus seis articulaciones tienen nombres en específico (Base, hombro, codo, muñeca 1, muñeca 2, muñeca 3), cada una tiene un rango de giro de ± 360°, a excepción de la muñeca tres que el rango de giro es ilimitado. La velocidad que pueden alcanzar las tres primeras articulaciones (Base, hombro y codo) es de ± 180°/s mientras que, las muñecas alcanzan velocidades de ±360°/s. Esto es de mucha importancia para conocer las limitaciones en cuestión de rotaciones del cobot.

Movimiento		
Movim. del eje del brazo robot.	Radio de acción	Velocidad máxima
Base	± 360°	± 180°/s
Hombro	± 360°	± 180°/s
Codo	± 360°	± 180°/s
Muñeca 1	± 360°	± 360°/s
Muñeca 2	± 360°	± 360°/s
Muñeca 3	Infinita	± 360°/s
Herramienta típica		1 m/s / 39,4 in/s

Figura 2.7: Movimientos del UR3

En la figura 2.8 se muestra la comunicación del robot UR3, ésta se puede realizar por medio de protocolos de comunicación como TCP/IP 100 Mbit con el estándar IEEE 802.3u, 100BASE-TX toma ethernet y MODBUS TCP. Actualmente, el UR3 que se encuentra en el CAP goza de una comunicación TCP/IP por medio de cable ethernet conectado al robot por medio de router cuatro puertos. Esto se logra haciendo uso de las opciones de configuración de red.

Figura 2.8: Pantalla del Teaching Pendant del UR3

A pesar de tener datos técnicos positivos pensando en el movimiento del robot es necesario realizar algunas adecuaciones al espacio de trabajo para el proyecto de grado. Además, por el uso de algunas herramientas como lo son, el gripper o pinza, cámara wrist y el sensor de torque y fuerza, se encuentran limitaciones adicionales por el cableado que usa con estas herramientas. Estas características y limitaciones se presentarán en la siguiente sección.

2.2.3. Dispositivos, adecuaciones y restricciones de movimiento

En esta sección se documentan los dispositivos, subsección 2.2.3.1 en el cual están las herramientas que hacen parte del robot (el gripper o pinza, cámara wrist y el sensor de torque y fuerza), los elementos de audio (diadema, micrófono) que son parte fundamental de la detección de los comandos de voz, también se presentan los elementos de comunicación que usa el robot, como cable de red y router. En la subsección 2.2.3.2 se documentan las limitaciones del robot o restricciones de movimiento como los grados de giro según las articulaciones, después de montar los sensores, la cámara y la pinza.

2.2.3.1. Dispositivos

El espacio de trabajo tiene algunas limitaciones ya incluidas por los sensores mencionados anteriormente. Los sensores con los que cuenta actualmente UR3 es un sensor FT 300-s, que se encarga de medir el torque y fuerza que ejerce el robot, una cámara Wrist y Gripper 2F-85. Las tres herramientas adicionadas al cobot son de la marca ROBOTIQ. Si bien no se usarán todos los sensores expuestos es necesario mencionarlos ya que hacen parte de la instrumentación de la que goza el UR3. La herramienta que será usada principalmente será el gripper, para el desarrollo del proyecto

de grado, sin este las tareas de *pick and place* (recoger y poner) planteados en las pruebas del experimento no se podrían realizar. Por otra parte, algunas de las principales características que poseen estas herramientas son las siguientes.

Comenzando con el sensor FT-300-s, sus especificaciones presentan que puede realizar mediciones en los 3 ejes cartesianos (X, Y, Z) en un rango de ± 300 N, una capacidad de sobrecarga del 500 % para los tres ejes mencionado. La deflexión en la máxima carga es de 0.01 mm y tiene una masa total de 440 g. Por otra parte, es inmune a la sensibilidad del ruido exterior y tiene una tasa de salida de datos de 100Hz. Por último, usa como protocolo de comunicación Modbus RTU/Data stream (RS-485). Ver figura 2.9.

Figura 2.9: Sensor de torque y fuerza, marca Robotiq

Ahora, en la figura 2.10 se encuentra la cámara Wrist cuenta con una masa de 160 g. La cámara posee algunas funciones programables como; enseñanza automatica de piezas y partes paramétricas, edición de bordes, color de objetos validación de autorización, modo avanzado y básico en el control de la cámara, permite varios parámetros como exposición, el foco, luz led, etc. Además, tiene un rango de operación de 0° a 50° C y beneficios cómo, programación en minutos de partes complejas, crear planos de trabajo solo con un click, creación automática de acción de picking con el centro de una posición entre otras muchas funciones.

Figura 2.10: Wrist camara, marca Robotiq

Ahora bien, las especificaciones del gripper Fig. 2.11 son; la fuerza con la que agarra es ajustable y tiene un rango entre 20 a 235N, la carga total que puede cargar la pinza es de 5 Kg, tiene una masa de 0,9 Kg y una resolución de posición de 0.4 mm, esto lo hace bastante preciso. La velocidad con la que cierra está en un rango de 20 a 150 mm/s también es ajustable. Por otra parte, respecto a la comunicación utiliza el protocolo Modbus RTU (RS-485) con IP45.

Figura 2.11: Gripper o pinza, marca Robotiq

Para las señales de audio se usaron en diferentes pruebas dos dispositivos diferentes una diadema Logitech G935 y un micrófono Yeti marca Blue muy conocido por su cancelación de ruido. A continuación se detallan las características de la diadema Logitech G935, cuanta con unas dimensiones de 19, 81 x 10, 67 x 23,11 centímetros (Alto, Ancho, Profundidad) y un peso de 0,72 kg. La respuesta en frecuencia es de 20 Hz, una sensibilidad de 93 dB y una impedancia de 39 Ohm eso en referencia a los audífonos. El micrófono tiene una respuesta de ancho de banda de 100 Hz. Figura 2.12. Estas diademas se usaron tanto para el dictado de comandos de voz como para la retroalimentación que provee la interfaz.

Figura 2.12: Diademas Logitech G Series G935

Por parte del micrófono Yeti fig. 2.13, tiene un consumo de voltaje de 5V a 150 mA, cuenta con una frecuencia de muestreo de 48kHz, una tasa de bits de 16-bit, para su cancelación de ruido cuenta con 3 cápsulas de condensador de 14mm. Además, cuenta con cuatro patrones polares o de escucha, los cuales son cardoide, bidireccional, omnidireccional y estéreo, el patrón usado en la pruebas fue el cardoide ya que este patrón permite que le micrófono solo escuche o detecte las ondas sonoras enfrente de el. Cabe resaltar que la interfaz se puede usar con cualquier tipo de micrófono o altavoces, la interfaz no discrimina estos equipos, si se recomienda el uso de equipos con disminución de ruido para ambientes de trabajo con esta dificultad. Este micrófono solo se usó para la captura de comandos de voz, para la retroalimentación se usó el altavoz del computador en el cual se corrió la interfaz de voz.

Figura 2.13: Micrófono Yeti marca Blue

2.2.3.2. Restricciones de movimiento

Por los sensores adicionados al UR3, que son el sensor de torque y fuerza, gripper y la cámara, estos llevan consigo un cableado para el transporte de la información que se recopila. Por tal motivo, los grados de libertad poseen limitaciones diferentes a las expuestas anteriormente en la sección 2.2.2, figura 2.7. Usando las funciones de movimiento libre del UR3 se procedió a determinar cuál sería los ángulos de rotación que tendría cada articulación. Se parte de una posición base que tiene los siguientes ángulos asignados a cada posición. En la figura 2.14, se puede apreciar los cables que se mencionan anteriormente.

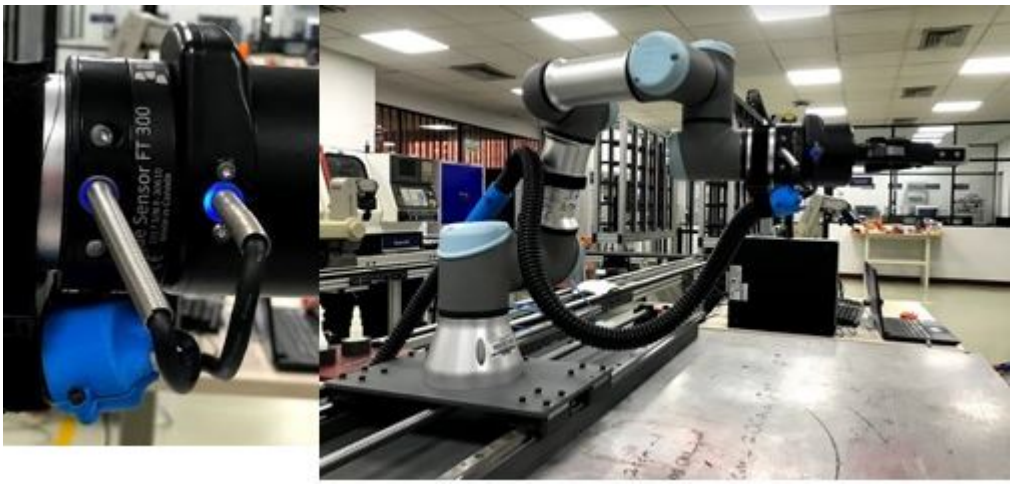


Figura 2.14: Cables presentes en todo le brazo robotico

La primera articulación, en la tabla 2.1 se aprecia como base, esta se giró en el sentido de las manecillas del reloj teniendo un desplazamiento aproximado de 228°, se repitió el procedimiento en el sentido contrario y el valor del desplazamiento fue de 231°. La medida del desplazamiento máximo se determinaba cuando el cable estaba ya en una posición donde podría resultar dañado o inducir al UR3 en un estado de bloqueo por sentir presión en sus articulaciones. Este mismo proceso se realizó para las demás articulaciones. Todos los datos recopilados están expuestos en la siguiente tabla.

Articulación	Radio de acción sin sensores	Radio de acción con sensores	Diferencia
Base	± 360°	± 230°	± 130°
Hombro	± 360°	± 200°	± 160°
Codo	± 360°	± 360°	± 0°
Muñeca 1	± 360°	± 270°	± 90°
Muñeca 2	± 360°	± 270°	± 90°
Muñeca 3	Infinito	± 270°	No medible

Cuadro 2.1: Tabla con los grados de rotación de cada articulación

2.2.4. Definición del ambiente de trabajo

El espacio de trabajo fig. 2.15 está conformado por las diferentes variables que componen el entorno, tales como la iluminación, ruido, base del robot y mesa donde se encuentra ubicado. También, en esta sección entra la definición del tablero de trabajo del UR3, los objetos que se van a usar para el experimento y sus características. La definición de estos permite tener unos límites de acción establecidos y claros.

Las condiciones de iluminación y ruido del ambiente del CAP son adecuadas para el desarrollo

de las actividades y pruebas del trabajo de grado. Inicialmente se pensaba que estas podrían afectar el desarrollo de alguna toma de señales. Sin embargo, el ruido presente en el ambiente no interviene tampoco en la captación de los comandos de audio. Adicionalmente, mirando las especificaciones de funcionamiento de los dispositivos que se usaron se encontraron ventajas ya que estos según sus datos técnicos trabajan en un rango de condiciones bastante amplio. La mesa de trabajo en la que se encuentra soportada el UR3 tiene como dimensiones 90 x 130 x 90 centímetros (Alto, Ancho, Profundidad). Debajo de ella se encuentra la caja de control del UR3 donde se puede apreciar las entradas y salidas de este. En el lado izquierdo está ubicada al UPS de seguridad para el UR3, para protegerlo contra descargas eléctricas (sobrevoltajes, descargas de tensión, etc.).

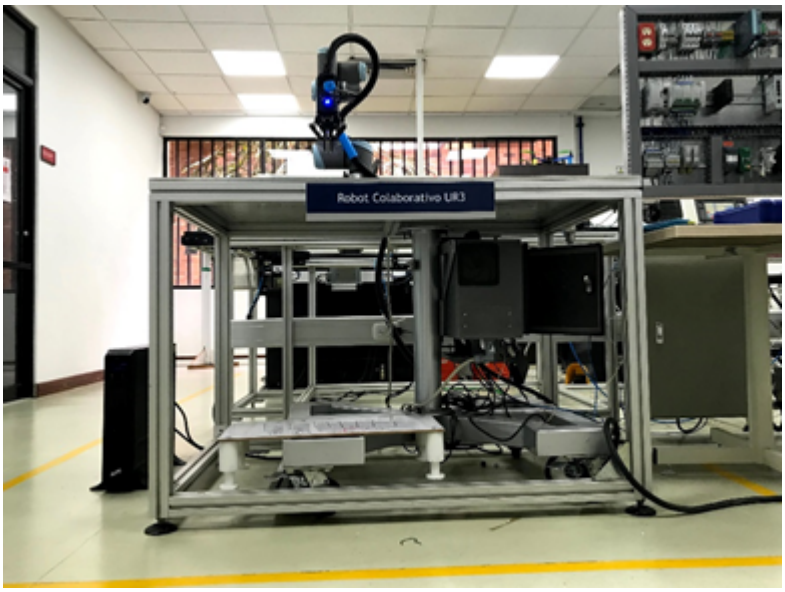


Figura 2.15: Plano de frente al Robot, se aprecia la caja de control o control box justo debajo de la mesa y esta la UPS de lado izquierdo de la mesa

La base del robot es una plataforma móvil que se puede ver en la figura 2.16, esta plataforma móvil hace parte de un proyecto diferente donde se pretende agregarle movimiento a la base del robot y obtener un séptimo grado de libertad. Por el momento, el proyecto no interviene de ninguna forma con el proyecto de grado y la base donde fue posicionado el UR3 no presenta ningún tipo de limitación para el desarrollo de actividades. Sin embargo, las variaciones de altura que presenta respecto a la mesa si es necesario tenerlas en cuenta. La base le agrega una altura de 7 centímetros de alto y un ancho de 130 centímetros, igual que la mesa.

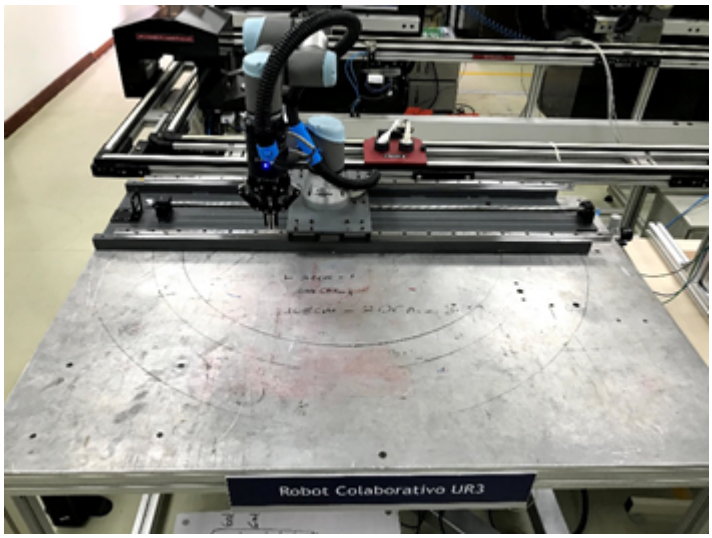


Figura 2.16: Plano desde arriba de la mesa del robot, se logra dimensionar la elevación del robot gracias a la plataforma movible en la que se encuentra posicionado

Por otra parte, se usaron cubos para desarrollar actividades del experimento, estos cubos serán todos del mismo tamaño con las mismas dimensiones que son (3 cm, 3 cm, 3 cm) (Alto, Ancho, Profundidad). En las figuras 2.17, 2.18, 2.19 se visualiza los cubos que se usaron en las pruebas. Estos cubos se imprimieron en las impresoras 3D del CAP y ya estaban disponibles desde antes de definir el proyecto de grado.

Figura 2.17: Cubo Amarillo

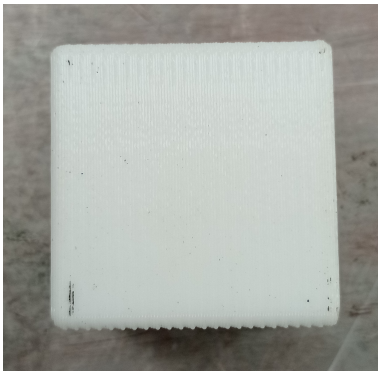


Figura 2.18: Cubo Blanco

Figura 2.19: Cubo Morado

La principal razón por la que se eligen este tipo de objetos es por la facilidad de agarre. Konrad Ahlin et al [54] explica que el agarre de objetos con un brazo robótico es una tarea difícil y muestra una alternativa de solución por medio del procesamiento de imágenes con una combinación de redes neuronales convolucionales (Por sus siglas en inglés Convolutional Neuronal Network CNN). Enten-

diendo que el enfoque del proyecto de grado no es solucionar un problema de agarre de diferentes objetos sino netamente de la interacción humano-robot.

A continuación, se define la posición inicial ver fig. 2.20, esta posición es común en las aplicaciones en robos ya que provee una posición referencia para el tipo de actividad o movimiento a realizar por parte del robot, y particularmente se define esta posición para el inicio de la interfaz ya que facilita los movimientos del robot en el área de trabajo a trabajar. Después de definir esta posición se inicia a determinar el área de trabajo.

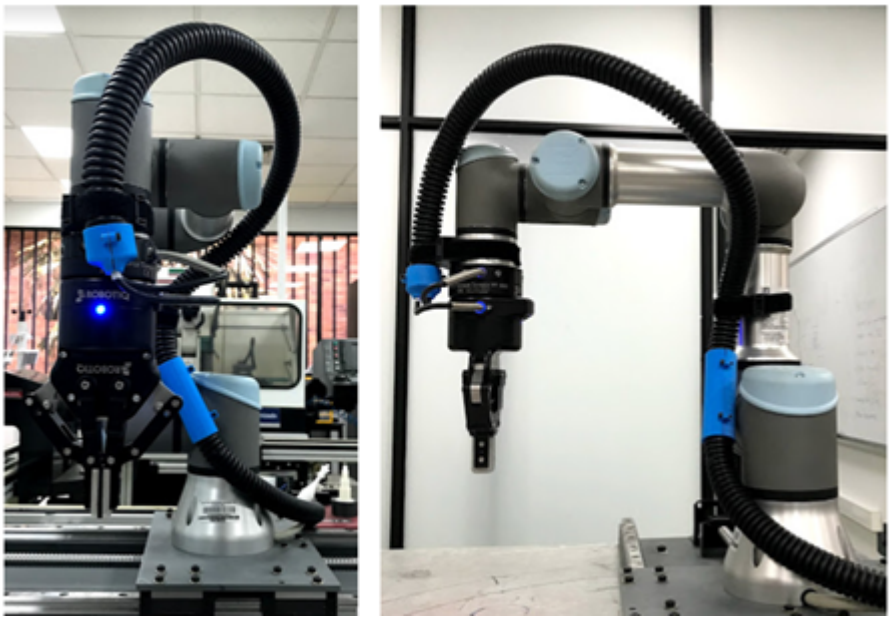


Figura 2.20: Posición inicial

Por último, el área de trabajo del UR3, es un recorte de tablero acrílico el cual tiene cuatro soportes con resortes los cuales protegen el tablero en caso de que le movimiento del robot sea bajar en el eje z, este movimiento denotaría una presión en el área de trabajo. Este es un espacio limitado que es el lugar donde el cobot tiene disponibilidad de realizar las actividades que se le exijan. No se tomarán acciones por fuera de este espacio, por ejemplo, si el usuario le exige al cobot colocar una de las fichas fuera del área este no realizará la tarea. También, aplica para situaciones donde se exija recoger una ficha por fuera. El área es la representada en la figura 2.21.

Figura 2.21: Área de trabajo

De forma general, el espacio de trabajo en conjunto con la mesa y el UR3 terminaría visualizándose como se ve en la figura 2.22. Se resalta que, el cuadrado amarillo que se encuentra frente a la mesa es el lugar donde deberá estar posicionado el usuario. Esta simulación del espacio de trabajo tiene el fin de mostrar al usuario el ambiente de trabajo más completo y no por separado para que el usuario tenga más comprensión acerca del ambiente de trabajo.

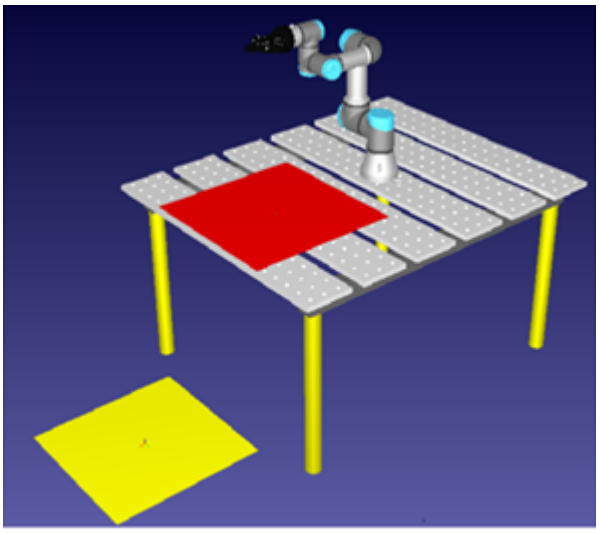


Figura 2.22: Simulación de espacio de trabajo en Robo Dk

2.3. Etapa de implementación

Para la etapa de implementación de la interfaz de voz se comparó las diferentes interfaces de voz que ya se han implementado en el robot UR3 para tener claro que tipo de diccionario de comandos

se realizaría para el robot UR3 del CAP de la universidad. En la sección 2.3.1 se documenta el diccionario que se definió para la interfaz de voz y se muestra que tipo de acción corresponde a cada comando de voz. Luego de definir el diccionario de comando se procedió a decidir que tipo de lenguaje de programación usar, una vez elegido se procedió a instalar las librerías de las cuales hará uso la interfaz de voz, en la sección 2.3.2 se muestra esto y también se recomienda que pasos seguir para la correcta implementación de la interfaz en cualquier computador que se requiera.

2.3.1. Definición del diccionario de comandos

Las interfaces de voz implementadas en los robots usan diferentes tipos de comandos de acción o movimiento, no hay una tabla de comandos definida para una interfaz de voz. Por lo general se usan verbos de acción que impliquen el movimiento preciso, como por ejemplo: "**Ir/Mover/Subir/Abrir**", etc. Estos y más verbos de acción parecidos se definieron para la interfaz de voz, también se definieron sinónimos de estos ya que ofrecen variables en el tipo de oración para el comando completo de voz.

La definición de lo comandos de audio se buscó que fueran fácil de usar por los usuarios, así su interacción con el robot sería menos tediosa. Se procuró utilizar un lenguaje natural y evitar llegar a un lenguaje técnico, poco usado por las personas en su diario vivir. Sin embargo, realizar una caracterización del lenguaje termina siendo difícil ya que, este está influido por diferentes componentes como el social, étnico, económico, grado de estudio y creencias, por nombrar solo algunos. Se buscó que la interfaz de voz incluyera diferentes palabras para una misma acción, así el diccionario de comandos se vería más enriquecido ya que los usuarios podrían elegir que palabra usar o tener la facilidad de usar cualquier palabra para referirse a una solo acción en particular, potenciando así el uso de la interfaz ya que adecua múltiples opciones de comandos de voz.

2.3.1.1. Comandos de audio

Los comandos de audio fueron divididos en tres secciones dependiendo de su utilidad, comando de movimiento, comandos de dirección y comandos de detalles. Existen más de una palabra que va a denotar la misma acción y en lo posible se busca tener la mayor cantidad de sinónimos para evitar este tipo de limitaciones. Primero definimos los comandos más importantes los de movimiento, estos son los denominados verbos de acción, en la tabla 2.2 se pueden ver estos comandos que se han definido para la interfaz de voz. La referencia al movimiento del robot se hace determinando la posición del operario en frente del robot.

Número	Comandos de movimiento	Descripción	Ejemplo de uso
1	Ir/Moverse	Este comando permite moverse hacia alguna de las 4 direcciones dependiendo del complemento.	"**Ir** hacia la izquierda 15 centímetros" "**Moverse** hacia arriba 2 centímetros"
2	Bajar/ Descender / Baja	Este comando permite que el brazo robótico baje (-Z)	"**Baja** 5 centímetros" "**Desciende** 12 centímetros"
3	Subir/ Ascender/ Sube	Este comando permite que el brazo robótico suba (+Z)	"**Sube** 6 centímetros" "**Asciende** 10 centímetros"
4	Abrir gripper /pinza	Permite abrir el gripper del UR3	"**Abrir gripper**" "**Abrir pinza**"
5	Cerrar gripper /pinza	Permite cerrar el gripper del UR3	"**Cerrar gripper**" "**Cerrar pinza**"
6	Medio gripper /pinza	Permite abrir o cerrar -según el caso- el gripper del UR3 a la mitad	"Abrir **Medio** el gripper" "Cerrar **Medio** la pinza"
7	Volver a casa	Este comando te permite posicionar al UR3 en una posición "casa" que es predeterminada	"**Volver a casa**"
8	Apagar Robot	Este comando lleva al robot a la posición de descanso para luego apagarlo	"**Apagar** el **Robot**"
9	Salir de la interfaz	Este comando finaliza la interacción con la interfaz de voz	"**Salir de la interfaz**"

Cuadro 2.2: Tabla con los comandos de movimiento de la Interfaz de Voz

Ahora bien, se definen otras tablas que son los complementos para los comandos de movimiento. Algunos de estos comandos se pueden evidenciar en los ejemplos de la tabla 2.2 como lo son "**Izquierda**", "**10 centímetros**", "**45 grados**" entre otros. En las siguientes tablas se podrán apreciar estos comandos de audio. Primero se muestra en la figura 2.3 los comandos de dirección que acompañarían al comando de movimiento "**Ir / Moverse**", estos comandos de dirección hacen referencia a las 6 direcciones que se tendrían de referencia para el movimiento, los movimientos hacia la derecha, la izquierda, arriba y abajo están referenciados por el usuario, es decir que el robot al decirle: "**Ir hacia la derecha**", el robot se moverá hacia la derecha del usuario. En cambio los movimientos hacia adelante y hacia atrás están referenciados con el robot, es decir que si al robot se le dice: "**Ir hacia atrás**", el robot acercará el efector final o pinza a su cuerpo.

Número	Comandos de dirección	Descripción	Ejemplo de uso
1	Derecha	Este comando hace mover al robot hacia la derecha del operario.	"Ir hacia la **derecha** 15 centímetros"
2	Izquierda	Este comando hace mover al robot hacia la izquierda del operario.	"Ir hacia la **izquierda** 15 centímetros"
3	Abajo	Este comando permite que el brazo robótico baje (-Z)	"Moverse hacia **abajo** 5 centímetros"
4	Arriba	Este comando permite que el brazo robótico suba (+Z)	"Ir hacia **arriba** 6 centímetros"
5	Atrás	Este comando te permite mover al UR3 hacia atrás con respecto a el	"Moverse hacia **atrás** 2 cm"
6	Adelante/ Delante/ Frente	Este comando mueve el robot hacia adelante	"**Adelante** 1 cm" "Ir hacia **Delante** 3 cm" "Moverse al **Frente** 1 cm"

Cuadro 2.3: Tabla con los comandos de dirección de la Interfaz de Voz

Ahora bien, para los comandos de detalles fig. 2.4 se define la cantidad de centímetros a moverse definido anteriormente el movimiento, sin este comando el robot no se moverá, si este comando falta la interfaz responderá al usuario que hace falta la cantidad de centímetros a moverse. En algún momento se pensó en realizar un movimiento definido de 1 centímetro si la cantidad de centímetros no estaba definida en el comando de voz del usuario, pero se descartó ya que se perdería el control en la actividad a realizar, ya que si cierta actividad ya tiene definida la cantidad de centímetros a moverse por el robot y este se mueve un centímetro más se perdería la precisión de la tarea.

Número	Comandos de detalles	Descripción	Ejemplo de uso
1	# de centímetros	Este comando permite especificar la cantidad de cm a mover en el robot	"Ir hacia la izquierda **15 centímetros**"

Cuadro 2.4: Tabla con los comandos de detalles de la Interfaz de Voz

Cada comando de voz tiene un feedback o retroalimentación del dialogo, el cual será una descripción de lo que esté haciendo o lo que hizo el robot, ejemplo "Me moví 5 centímetros a la derecha", "Estoy abriendo el gripper o pinza", este feedback se dará por medio de los altavoces que estén predeterminados en el computador que este corriendo el programa (**.py**).

2.3.2. Implementación del código

Para la implementación de la interfaz se usó el lenguaje de programación Python(Py), ya que este representa ventajas significativas al usar librerías ya definidas en el lenguaje. La versión de Py que se usó fue **3.9.5**, para la instalación de las librerías adicionales que necesita la interfaz se usó el administrador de paquetes de Py *pip* en la versión **21.1.2**.

Las librerías usadas de **py** son, *speech_ recognition* , para la etapa de reconocimiento del habla, esta librería se instala con la siguiente orden en el cmd: *pip install SpeechRecognition*. La documentación de esta librería se encuentra en el siguiente link: https://pypi.org/project/SpeechRecognition/. Para la etapa de síntesis de voz se usa *pyttsx3*, esta librería se instala con la siguiente orden en el cmd: *pip install pyttsx3*. La documentación de esta librería se encuentra en el siguiente link: https://pypi.org/project/pyttsx3/.

`import speech_recognition as sr`, se usa para importar la librería de reconocimiento del habla y `import pyttsx3` para la librería de síntesis de voz.

Ahora bien estas dos librerías hacen parte de un sistema de lenguaje hablado(SDS) que se mostró en el marco teórico, el entendimiento del habla, manejador de dialogo y la generación del lenguaje estarían implícitas en las sentencias o lineas de código en el programa implementado. Para la etapa del entendimiento del lenguaje se toma cuando se captura la entrada de voz en una variable y se toman decisiones según la entrada. El manejador de dialogo se materializa en los condicionales para la realización de acciones según corresponda y para la etapa de generación del lenguaje hace parte del código dentro de los condicionales ya nombrados.

Para poder usar el micrófono para el reconocimiento del habla en python hace falta usar la librería *PyAudio* cuya documentación de esta se encuentra en le siguiente enlace: https://pypi.org/project/PyAudio/. En el repositorio donde se encuentra alojado todos los programas realizados para esta interfaz de voz se encuentran más detalles de instalación es esta librería importante para el dispositivo de adquisición de señal de la entrada. El link de este repositorio se encuentra al finalizar la sección.

Se usa otra librería de **py**, esta para realizar notificaciones, esta librería se instala con la siguiente orden en el cmd: *pip install win10toast*. La documentación de esta librería se encuentra en el siguiente link: https://pypi.org/project/win10toast/. De *win10toast* se importa **ToastNotifier**, así `from win10toast import ToastNotifier`. Luego, se usa está función para notificar al usuario al iniciar la interacción de la interfaz. Esta librería es natural del sistema operativo *Windows*, por tanto si se quiere implementar la interfaz en otros sistemas operativos, se debe instalar otras dependencias o de ser necesario quitar las notificaciones de la interfaz, excluyendo del código lineas que contengan o usen la librería.

Para el control del UR3 se usa comunicación por socket, el cual es un proceso servidor-cliente

basado en un conjunto de protocolos TCP/IP, para establecer comunicación con el robot se debe saber la IP del robot y el puerto de comunicación disponible para la conexión, en nuestro caso en la fig. 2.23 se pueden ver los requerimientos de comunicación.

```
HOST = "192.168.0.100" # ip del robot
PORT = 30002 # puerto en el que el robot recibe
```

Figura 2.23: Definición de HOST y PORT

El socket que se crea para la comunicación con el robot se usa para enviarle al UR3 comandos del lenguaje de programación URScript, que es el lenguaje creado por Universal Robots para controlar el robot UR3. Se debe hacer este proceso ya que los comandos del lenguaje URScript no los puede compilar un computador, entonces se debe hacer un programa en **py** para luego hacer comunicación vía socket y enviar los comandos de movimiento tipo URScript. Los comandos del lenguaje URScript se pueden encontrar en el siguiente enlace: https://s3-eu-west-1.amazonaws.com/ur-support-site/32554/scriptManual-3.5.4.pdf

Algo importante a destacar es que los URScript son comandos que solo competen a funcionalidades del robot, movimientos, giros, manejo libre o freedrive, etc. Es decir, que para usar los dispositivos adicionales, el gripper, la cámara o el sensor de torque y fuerza por medio de código, no es posible con los URScript ya que estos dispositivos o accesorios son adicionales al robot y son marca Robotiq, con esto denotamos que estos accesorios son externos a Universal Robots y por ende no son compatibles con URScript. Por consiguiente, para el manejo del gripper o la pinza se usan 4 archivos para las acciones de activar, abrir, cerrar y posición media del gripper, en estos se encuentran diferentes instrucciones que sirven para comandar el gripper, estos archivos se envían vía socket en forma binaria.

Por ultimo y no menos importante, además de la cancelación de ruido que ya proporcionan los micrófonos ya sea el micrófono de la diadema o el micrófono Yeti la librería de *speech_recognition* también permite un ajuste de ruido ambiente el cual usa un tiempo antes de iniciar a reconocer el habla, este tiempo se puede definir en un valor determinado, un segundo o menos de ser necesario. Esto es importante si se desea usar la interfaz de voz para tareas donde el robot este en ambientes un poco más ruidosos que el CAP de la universidad. La interfaz se encuantra alojada en el siguiente repositorio: https://github.com/SebastianBlandon/UR3_VOICE_INTERFACE.git.

2.4. Etapa de evaluación

El desarrollo de la etapa de evaluación se fraccionó en tres secciones, en la primera sección la 2.4.1 se diseña de un protocolo de pruebas, en el cual se muestran las pruebas que se hicieron y

que tipo de tarea desarrolló el robot UR3. En la segunda sección la 2.4.2 se definen las métricas cuantitativas para la evaluación de las pruebas que realizó cada usuario. Por ultimo, en la tercera sección la 2.4.3 se muestran las métricas cualitativas para la evaluación de las pruebas que realizaron los usuarios que colaboraron.

2.4.1. Diseño de un protocolo de pruebas

Para el diseño del protocolo de pruebas se plantea realizar con el robot una tarea de ensamble como las que se ven en las lineas industriales en las cuales integran robots, en áreas como la automotriz, almacenamiento, distribución, etc. En general, una tarea típica de ensamble que use robots implica operar con dos o más objetos. Cada parte es un subconjunto del ensamble. El objetivo del ensamblaje es calcular un orden de operaciones que reúne partes individuales para que aparezca un nuevo producto o actividad prevista, como ordenar y apilar objetos. Algunos ejemplos de tareas de ensamblaje típicos son:

- Clavija en el agujero (Peg-in-hole): una pinza robótica agarra la clavija y la inserta en un agujero.
- Trayectoria irregular (Irregular trajectory): el robot sujeta un objeto que deposita material sobre otro objeto a lo largo de una trayectoria irregular.
- Deslizar en la ranura (Slide-in-the-groove): un robot inserta un perno que se ajusta dentro de una ranura y desliza el perno a la posición deseada donde se va a quedar fijado.
- Atornillar (Bolt-screwing): un robot atornilla un tornillo en un material de propiedades desconocidas.
- Ensamblaje de silla (Chair-assembly): un robot integra las piezas de la silla junto con un sujetador.
- Conexión de tubería (Pipe-connection): un robot recoge y coloca dos tuercas de unión en un tubo.
- Agarre y ubique (Pick-and-place): un robot recoge un objeto en un lugar de inicio y lo deposita o coloca en un lugar o una posición final .

La tarea de ensamble elegida para la realización de las prueba fue hacer un *pick and place*, cada pruebas puede contener una o varias de estas tareas dependiendo de la prueba se hará uno, dos o tres de estas tareas, así realizando una actividad completa de movimiento del robot y los objetos con los que se realizan las pruebas.

TAREA:
Las pruebas constaran de llevar a cabo una actividad de *"Pick and Place"*. Esta consiste en tomar

un objeto de una posición A, fig. 2.28 y colocarla en una posición B, fig 2.29. El usuario deberá hacer las diferentes pruebas manejando el robot mediante comandos del Teaching Pendant(TP) y por medio de comandos de voz, en ambos casos se le explicará al usuario como manejar el TP y como interactuar con la interfaz de voz. Dentro de este proceso se tomarán en cuenta diferentes parámetros que serán analizados como lo son la precisión, la velocidad, el número de intentos y el número de instrucciones dadas. Con esto, además de lograr comprobar la hipótesis propuesta también se busca medir el rendimiento y capacidades de la interfaz de voz desarrollada en el proyecto de grado.

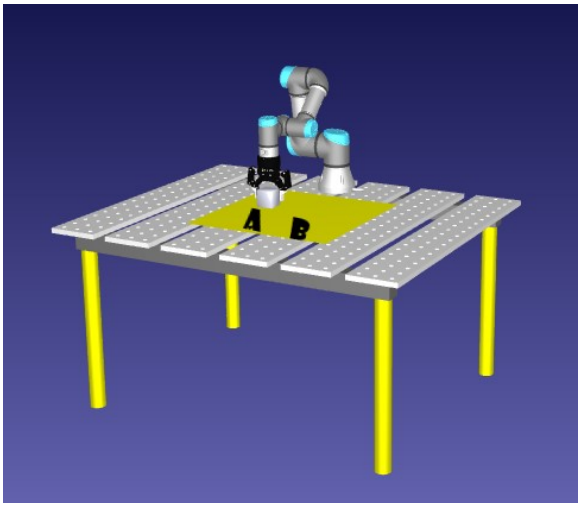


Figura 2.24: El robot levanta el cubo (*Pick*)...

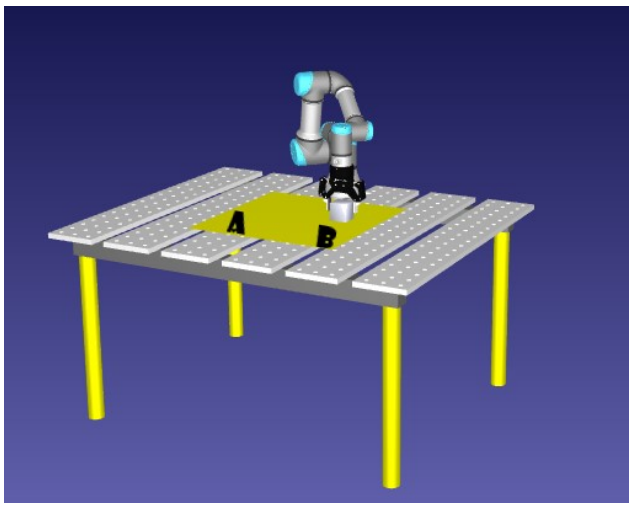


Figura 2.25: El robot pone el cubo (*Place*)

PRUEBA 1

En esta primera prueba el usuario deberá programar el movimiento del robot mediante el TP y también le deberá pedir por medio de los comandos de voz al UR3 que mueva el cubo de color **amarillo** al punto predeterminado **A**. En la figura 2.26 se puede ver que el cubo amarillo está situado es una posición inicial predeterminada y se debe mover éste a la posición **A** que está posicionada diagonalmente a la posición inicial.

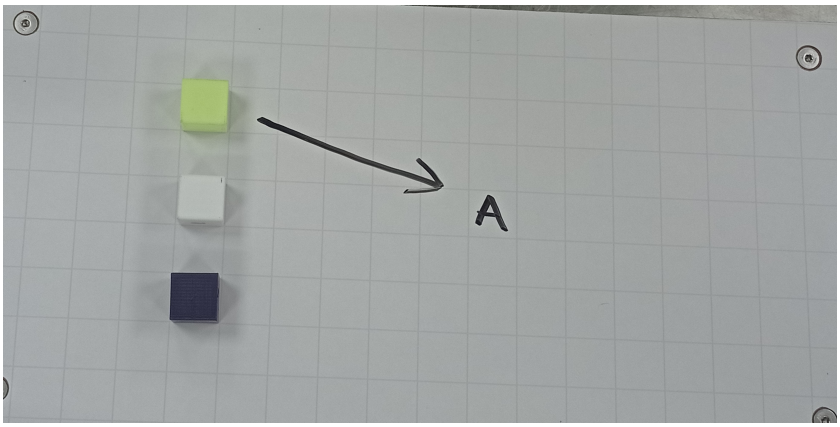


Figura 2.26: Ilustración prueba 1

PRUEBA 2

En la segunda prueba el usuario deberá pedirle al UR3 por medio del TP y la interfaz de voz que mueva dos objetos de diferentes colores a posiciones A y B que son puntos predeterminados en la práctica. En la figura 2.27 se evidencian las posiciones iniciales y finales a las cuales se deben llevar los cubos **blanco** y **morado**.

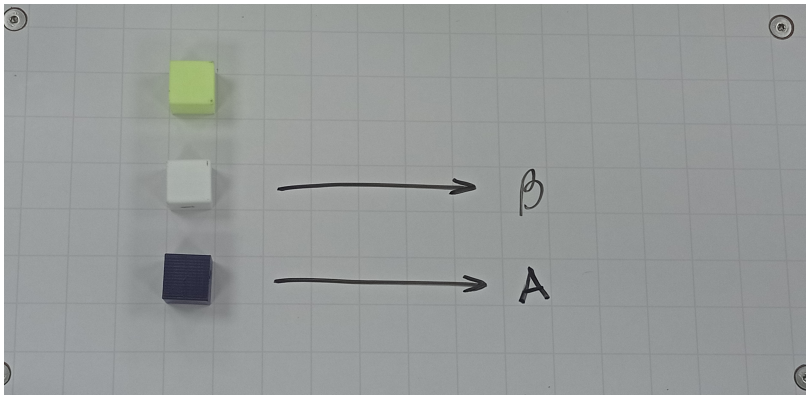


Figura 2.27: Ilustración prueba 2

Las pruebas 1 y 2 tienen asociadas unas reglas que permitirán definir si estas fueron hechas correctamente o si, por el contrario, deben ser repetidas.

Reglas:

1. Puede asignarle pequeñas tareas que terminen cumpliendo la general que es cambio de posición espacial de la ficha elegida por el usuario.
2. Para aceptar que la prueba fue realizada correctamente la ficha se debe encontrar como máximo a 4 cm de los puntos.

3. Debe mover a los puntos solo la fichas predispuestas para el movimiento, si mueve una ficha diferente a estas, aunque cumpla las anteriores reglas no será tomado en cuenta como correcto.

PRUEBA 3

Esta es la última prueba donde se demanda realizar un ejercicio que demanda más tiempo que los demás. En esta prueba se debe asignarle al UR3 que intercambie de posición las tres figuras, siguendo la posicion inicial con se muestra en la fig. 2.28 y se debe terminar la prueba cuando los tres cubos este posicionados como en la figura 2.29.

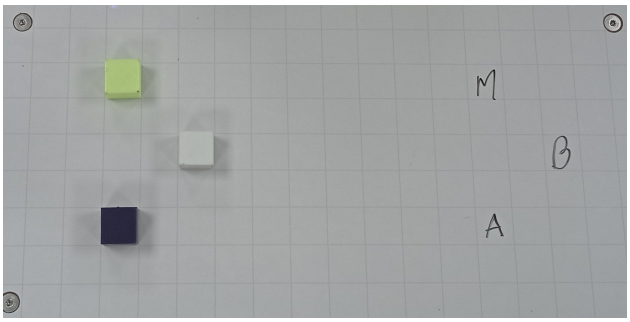


Figura 2.28: Posición inicial de las fichas para la prueba 3

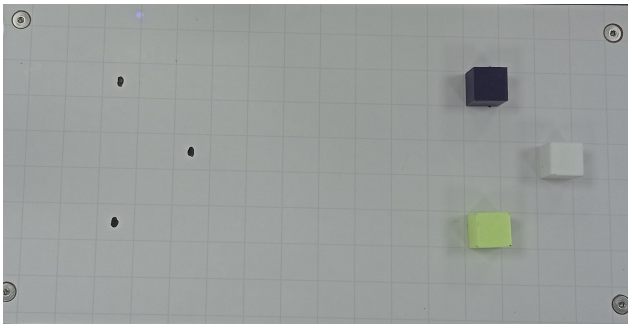


Figura 2.29: Posición final de las fichas para la prueba 3

Reglas:

1. Puede asignarle pequeñas tareas que terminen cumpliendo la general que es cambio de posición espacial de las fichas elegidas por el usuario.
2. Para aceptar que la prueba fue realizada correctamente, las fichas se debe encontrar como máximo a 4 cm del punto central donde se ha marcado el sitio de *place* de cada una de las fichas.
3. Debe intercambiar las fichas elegidas tal cual como en las imágenes, si intercambia diferente las fichas automáticamente no será tomado como correcto.

2.4.2. Métricas de evaluación cuantitativas

El tiempo

El tiempo se medirá en tres momentos. Primero, en el "pick" cuánto tarda desde el momento que recibe la orden hasta el momento donde agarra el cubo. Segundo, el "place" que comienza desde el momento donde agarró el objeto hasta que termina colocándolo en la posición pedida por el usuario. Tercero, es la suma de ambos tiempos para tener el tiempo que se demoró en realizar toda la tarea exigida.

Número de intentos

Este es uno de los puntos más importantes de la medición, ya que de requerir más de un intento para realizar cualquiera de las tareas propuestas por el usuario la interacción humano-robot se vería afectada. Por ejemplo, si usted estuviera sosteniendo una conversación con un compañero de trabajo y usted tiene que repetir 3 o 4 veces el mensaje que desea transmitir la comunicación que existe entre ambos sería poco efectiva. Lo mismo ocurre en el caso del experimento, en el caso ideal se debería completar la tarea propuesta por el usuario en un solo intento.

Número de instrucciones

Para finalizar, el número de instrucciones evaluará aspectos como el número de detalles o de información que el usuario debió propiciarle al UR3 para completar la tarea. Por ejemplo, si el UR3 se le pide que mueva un cubo a un punto especifico con una cantidad de centímetros definidos se espera que lo haga sin tener que adicionar detalles como un centímetro más hacía la izquierda, derecha, arriba, etc. Por otra parte, este aspecto también influye en la situación presentada del número de intentos. Si un usuario debe entregar muchos detalles al robot para cumplir adecuadamente la tarea también se vería afectada y no sería muy efectiva. Por consiguiente, es uno de los aspectos principales a evaluar.

En el momento de llevar a cabo el experimento se dividirá en dos partes. La primera parte se llevará a cabo esta actividad solo haciendo uso del TP. En la segunda parte se podrán usar la interfaz de voz, como lo antes dicho. Se espera después de realizar el experimento que el hacer uso de los comandos de audio hayan presentado un mejor rendimiento. Ambos casos deberán ser evaluados con la misma métrica para ser comparados adecuadamente. Cuando se menciona "métrica" solo es el conjunto de variables mencionadas anteriormente dentro de una tabla, esto permite presentar la información de manera más ordenada y entendible. Ver tabla 2.5.

Métrica			
Parámetros	**Medibles**		
Tiempo	¿Cuánto se demoró haciendo toda la tarea?	¿Cuánto se demoró en el pick?	¿Cuánto se demoró en el place?
Número de intentos	Número de intento para el pick	Número de intentos para el place	Número total de intentos
Número de instrucciones	Número de instrucciones para el pick	Número de instrucciones para el place	Número total de instrucciones

Cuadro 2.5: Tabla con las métricas de evaluación cuantitativas

2.4.3. Métricas de evaluación cualitativas

La precisión

Se habla de precisión cualitativa ya que se evaluará una precisión en los comando al momento de realizar la acción, como recoger el objeto del lugar correcto, también, dejar el objeto en el lugar indicado. Este tipo de evaluación cualitativa se denomina realizando análisis de la tarea realizada por el robot después de recibir el comando. Para esta evaluación de califica usando una escala likert y se asigna un valor a cada cualidad que se defina a evaluar.

En la tabla 2.6 se muestra un ejemplo de como fue medible la precisión con tres tipos de características, las distancia el punto de referencia en recoger y colocar la ficha, el agarre de la pinza y cuando suelta la pinza y la recogida de la ficha correcta y posicionarla en el lugar correcto, estas características se califican con valores de *Muy malo*, *Malo*, *Bien*, *Muy Bien* y *Excelente* definidas en una escala likert. El agarre se midió dependiendo de como la pinza recogió el objeto, si lo tomó bien en el centro de este o si lo tomó de una esquina y demás.

Métrica			
Parámetros	**Medibles**		
Precisión (Pick)	Distancia del punto de referencia de donde fue recogida la ficha	¿Agarró correctamente el objeto?	¿Recogió la ficha correcta?
Precisión (Place)	Distancia del punto de referencia a donde fue colocada la ficha	¿Soltó correctamente el objeto?	¿Dejó la ficha correcta?

Cuadro 2.6: Tabla con las métricas de evaluación cualitativas

Capítulo 3

Análisis de resultados

Para la realización de las pruebas se contó con algunos inconvenientes que fueron difíciles de manejar como lo fue la poca afluencia de personas en la universidad por causa de la pandemia propiciada por la Covid-19, adicional a esto en las fechas que se habían previsto las pruebas se presentó el paro nacional, por tal razón para las pruebas se contó con tan solo ocho personas. El número de personas hubiese sido mayor pero algunas personas manifestaron no colaborar con las pruebas ya que se sentían incapaces para aprender el lenguaje del robot ya que una parte de la prueba se realizaba aprendiendo a programar el robot mediante Teaching Pendant(TP), una persona por falta de tiempo solo pudo estar en la prueba unos 10 minutos nada más (dejando sin completar la prueba), y cada prueba en sí puede variar su duración de 20 a 30 minutos (Agregando el tiempo de explicación del modo de uso y manejo del TP – unos 10 minutos), y al ser tres pruebas y algunas pruebas demandan más tiempo (la 2 y la 3) por tener que mover más piezas, entonces el tiempo total de las 3 pruebas varia de 1 hora y 10 minutos a 1 hora y 40 minutos, por lo cual muchos de los usuarios que ayudaron con las pruebas no disponían de todo ese tiempo por tal razón solo una persona dispuso de su tiempo para realizar todas las pruebas.

A raíz de haber contado con pocos usuarios y con pocas pruebas por usuario, se determinará un análisis general de las pruebas luego de documentar y analizar uno a uno los resultados de las pruebas de cada usuario. Pensando en preservar la identidad de cada uno de los usuarios se usará nombres genéricos para cada uno de los usuarios que colaboraron con las pruebas. Se realizaron preguntas preliminares acerca del conocimiento previo en el manejo de robots y en el manejo de interfaces de voz o asistentes de voz. Se presentan algunas características de los usuarios como sexo, edad, ocupación y carrera, para mostrar que las pruebas se realizaron con usuarios de diferentes sexos, edades y con diferentes niveles de educación.

Para que la realización de los experimentos o pruebas finales fuese la más adecuada y cómoda para el usuario se realizaron muchas pruebas preliminares que sirvieron para realizar todas las adecuaciones al código para que la interfaz de voz funcionara correctamente y que su objetivo fuese tener una conversación natural con el usuario. Para esto fue fundamental haber definido muy bien las palabras a usar en los comandos de voz, ya que se notó en las pruebas realizadas la flexibilidad que ofrecían estos comandos y así no depender de una única oración para cada movimiento asociado a cada comando de voz. Solo bastó explicarle a cada usuario que tipo de comandos estaban definidos y cada usuario naturalmente ya formaba la oración que sería su comando de voz.

Para tener una buena experiencia con la interfaz se recomienda hablar claro y fuerte al micrófono

si este está un poco alejado, si se tiene el micrófono cerca ya sea el de una diadema, audífonos o un micrófono externo, se tendrá mejor experiencia de interacción ya que en algunos ambientes ruidosos para la interfaz es complicado reducir al 100 % estos fragmentos de la señal que no hacen parte de ella, puesto que ruido es todo aquello que no haga parte de la voz (en este caso particular), ruido sería un sonido constante que interrumpa el comando de voz, o puede ser otra persona hablando que diga palabras diferentes a las descritas en el diccionario de comando, en este caso la interfaz no reconocería el comando de voz. Por tal razón, particularmente para algunos usuarios la interfaz demoraba haciendo el reconocimiento de voz y detectando el comando, una vez detectado el comando de voz, la demora el en movimiento del robot se atribuyó a demoras en la comunicación entre el robot y el computador.

En algunas pruebas los usuarios decían el comando de voz bien pero no lo decían en el momento preciso en el cual el programa estaba recibiendo la información, lo decían a destiempo. Es decir, el programa tiene definido un tiempo en el cual se activa la adquisición de los comandos de voz por medio del micrófono definido para dicha adquisición de la señal, entonces el usuario dispone de un tiempo especifico para decir el comando de voz, este tiempo inicia cuando el programa notifica al usuario que el micrófono está activado. También se presentaron casos en los cuales el usuario hacía pausas, es este caso el programa cierra la adquisición de información por parte del micrófono y empieza a procesar lo que recibió, entonces los comandos de voz se enviaban incompletos y la tarea no se realiza.

En la tabla 3.1 se evidencian los resultados de la variable cuantitativa tiempo de los ocho usuarios usando el Teaching Pendant, realizando la tarea *pick* y *place*. En este resumen de los resultados se puede notar que el tiempo mínimo en el *pick* fue 3 minutos, el tiempo mínimo en el *place* fue 5 minutos y el tiempo mínimo de toda la prueba fue 8 minutos. Por parte del tiempo máximo, en el *pick* fue 8 minutos, en el *place* fue 9 minutos y en toda la prueba fue 16 minutos. Por último, se calculó el promedio de cada tiempo, por parte del *pick* se promedió un tiempo de 6 minutos con una desviación estándar del 2,07 % (no hay mucha dispersión entre los datos), por parte del *place* se promedió un tiempo de 6 minutos 37 segundos con una desviación de 1,3 % y para toda la prueba se promedió un tiempo de 12 minutos y 37 segundos con una desviación de 3 % (hay un poco más de dispersión en los datos).

Resultados Teaching Pendant			
Usuarios	**Tiempo (Minutos)**		
	Pick	Place	Toda la prueba
Usuario 1	8	7	15
Usuario 2	4	7	11
Usuario 3	8	7	15
Usuario 4	6	6	12
Usuario 5	4	5	9
Usuario 6	7	9	16
Usuario 7	8	7	15
Usuario 8	3	5	8
Mínimo	3	5	8
Máximo	8	9	16
Promedio	6	6,625	12,625
Desviación estándar	2,070196678	1,302470181	3,067688753

Cuadro 3.1: Tabla con el resumen de tiempos de los 8 usuarios usando el TP, se añade valor Mínimo, Máximo, Promedio y Desviación estándar

En la tabla 3.2 se presentan los resultados de la variable cuantitativa tiempo (de la tarea de *pick* y *place*) de los ocho usuarios usando la interfaz de voz. En este resumen de los resultados se puede notar que el tiempo mínimo en el *pick* fue 2 minutos, el tiempo mínimo en el *place* fue 2 minutos y el tiempo mínimo de toda la prueba fue 5 minutos. Por parte del tiempo máximo, en el *pick* fue 6 minutos, en el *place* fue 8 minutos y en toda la prueba fue 13 minutos. Por último, se calculó el promedio de cada tiempo, por parte del *pick* se promedió un tiempo de 4 minutos con una desviación estándar del 1,19 %, por parte del *place* se promedió un tiempo de 4 minutos con 37 segundos con una desviación de 1,99 % y para toda la prueba se promedió un tiempo de 8 minutos y 37 segundos con una desviación de 3 % (hay un poco más de dispersión en los datos).

Resultados Interfaz de Voz			
Usuarios	**Tiempo (Minutos)**		
	Pick	Place	Toda la prueba
Usuario 1	6	7	13
Usuario 2	3	2	5
Usuario 3	4	4	8
Usuario 4	4	4	8
Usuario 5	4	5	9
Usuario 6	4	4	8
Usuario 7	5	8	13
Usuario 8	2	3	5
Mínimo	2	2	5
Máximo	6	8	13
Promedio	4	4,625	8,625
Desviación estándar	1,195228609	1,995530721	3,067688753

Cuadro 3.2: Tabla con el resumen de tiempos de los 8 usuarios usando la interfaz de voz, se añade valor Mínimo, Máximo, Promedio y Desviación estándar

Se notó por los resultados de las pruebas de las tablas 3.1 y 3.2 que con la interfaz de voz se obtienen mejores tiempos en toda la prueba, obteniendo mínimos tiempos menores al que se dan con el uso del TP. También, los máximos tiempos son menores a los que se dan usando el TP, los promedios de tiempos también son inferiores en toda la prueba haciendo uso de la interfaz de voz. Incluso, las dispersiones de los datos estudiados son menores con el uso de la interfaz de voz ya que las desviaciones de los datos son inferiores comparativamente cuando se da el uso del TP.

Para la variable cuantitativa número de intentos se pudo observar que no hubo fallos en ninguna de las interfaces ya que solo fue necesario un intento para poder tener control de movimiento del robot. Tal vez, se sabia que al controlar al robot por medio el Teaching pendant esta situación se daría ya que el modo de programación es desarrollado por la empresa que fabrica el robot, por tal razón no debería tener fallos. Lo que se buscaba con este parámetro era encontrar las falencias de la interfaz de voz pero al tener estos resultados se puede afirmar que con un buen uso de la interfaz esta no debería fallar en ningún momento. Si hay algún fallo en la comunicación se debe revisar la configuración del router o si las peticiones en el robot se ha interrumpido por algún tipo de error en un movimiento. Por ejemplo, cuando encuentra una singularidad en la trayectoria a moverse o si el robot se encuentra parado por algún tipo de estado de espera o alarma, como cuando se presiona el botón de parada.

Para el número de instrucciones se puede ver en la tabla 3.3 los resultados haciendo uso del TP del número de instrucciones en el *pick*, *place* y en la cantidad total. Se calculó la cantidad mínima de instrucciones y para el *pick* y el *place* fue la misma 3 instrucciones, el minimo total de instrucciones

fue 6. Por parte de las maxima cantidad de instrucciones usadas fue de 6 para el *pick*, 8 para el *place* y en total una cantidad de 13 instrucciones. En promedio se usaron 4,4 instrucciones para el *pick* con una desviación del 1,3 %, 5,5 instrucciones para el *place* con una desviación del 1,5 % y en total de instrucciones se usaron en promedio 9,9 instrucciones con una desviación del 2,6 %.

Resultados Teaching Pendant			
Usuarios	**Número de instrucciones**		
	Pick	Place	Toda la prueba
Usuario 1	3	5	8
Usuario 2	5	8	13
Usuario 3	6	7	13
Usuario 4	6	6	12
Usuario 5	3	3	6
Usuario 6	4	5	9
Usuario 7	5	5	10
Usuario 8	3	5	8
Mínimo	3	3	6
Máximo	6	8	13
Promedio	4,375	5,5	9,875
Desviación estándar	1,302470181	1,511857892	2,587745848

Cuadro 3.3: Tabla con el resumen del número de instrucciones de los 8 usuarios usando el TP, se añade valor Mínimo, Máximo, Promedio y Desviación estándar

En la tabla 3.3 se muestran los resultados del número de instrucciones usadas haciendo las pruebas con la interfaz de voz, para el *pick*, *place* y en la cantidad total de instrucciones. Se calculó la cantidad mínima de instrucciones y para el *pick* y el *place* fue la misma 3 instrucciones, el minimo total de instrucciones fue 7. Por parte de las maxima cantidad de instrucciones usadas fue de 5 para el *pick*, 7 para el *place* y en total una cantidad de 12 instrucciones. En promedio se usaron 4,1 instrucciones para el *pick* con una desviación del 0,64 %, 4,2 instrucciones para el *place* con una desviación del 1,28 % y en total de instrucciones se usaron en promedio 8,4 instrucciones con una desviación del 1,68 %.

Resultados Interfaz de Voz			
Usuarios	**Número de instrucciones**		
	Pick	Place	Toda la prueba
Usuario 1	5	4	9
Usuario 2	4	3	7
Usuario 3	4	4	8
Usuario 4	4	3	7
Usuario 5	4	4	8
Usuario 6	4	5	9
Usuario 7	3	4	7
Usuario 8	5	7	12
Mínimo	3	3	7
Máximo	5	7	12
Promedio	4,125	4,25	8,375
Desviación estándar	0,640869944	1,281739889	1,685018016

Cuadro 3.4: Tabla con el resumen del número de instrucciones de los 8 usuarios usando la interfaz de voz, se añade valor Mínimo, Máximo, Promedio y Desviación estándar

Se nota en las tablas 3.3 y 3.4 que los resultados en comparación de las dos interfaces (gráfica y de voz) son muy parecidos, difieren un poco en la dispersión de los datos ya que se obtiene mejores desviaciones de los datos con la interfaz de voz. En general, los resultados de la interfaz de voz están un poco más bajos eso es bueno pensando que se usan la misma cantidad de instrucciones en mucho menos tiempo con la interfaz de voz, dando ventajas el momento de tener poco tiempo para poder interactuar con el robot. Este acierto hace pensar que si las pruebas se hubiesen hecho solo con la interfaz de voz muchas más personas de las cuales no pudieron colaborar con las pruebas por falta de tiempo hubiesen podido interactuar con la interfaz de voz. No solo se piensa en el tiempo de la persona que usará el robot, sino también hay ventajas en tareas que se deban realizar rápido por tener algún riesgo si hay demora, por ejemplo, el desarme de una bomba o inclusive una operación a corazón abierto, tareas que si hay demoras se pierden vidas.

A continuación, se documenta la variable cualitativa precisión en el *pick* y en el *place*. Primero se logra ver en la gráfica de barras de la fig. 3.1 los resultados de la precisión en el *pick* para la distancia del punto de referencia de donde fue recogido el objeto haciendo uso del TP, se puede ver que el 50 % de los usuarios, es decir, 4 de ellos recibieron una calificación de **bien** o 3 según la escala, dos de ellos un 25 % obtuvieron calificación de **muy bien** y los últimos dos 25 % una calificación de **excelente**. En comparación con la gráfica de la fig. 3.2 donde se encuentran los resultados de la precisión en el *pick* para la distancia del punto de referencia de donde fue recogido el objeto haciendo uso de la interfaz de voz, 3 de los usuarios un 37,5 % obtuvieron una calificación de **muy bien** y los restantes 5 el 62,5 % de la población obtuvieron una calificación de **excelente**. Con esto,

se puede evidenciar que en este aspecto fue más preciso en el *pick* el robot con la interfaz de voz.

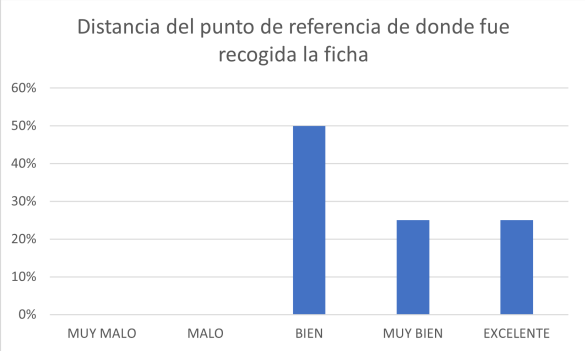


Figura 3.1: Resumen de resultados de la variable cualitativa precisión en el pick en la distancia al punto de referencia de los 8 usuarios usando el Teaching Pendant

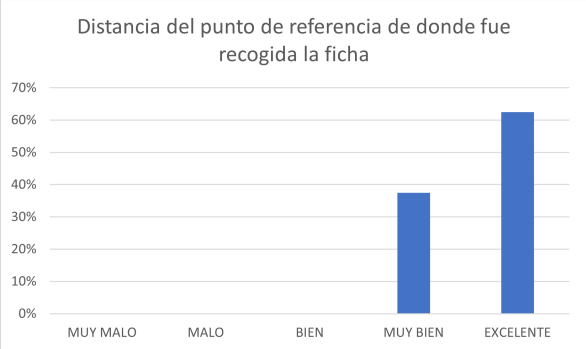


Figura 3.2: Resumen de resultados de la variable cualitativa precisión en el pick en la distancia al punto de referencia de los 8 usuarios usando la interfaz de voz

Se puede ver en la gráfica de barras de la fig. 3.3 los resultados de la precisión en el *pick* para agarre correcto del objeto haciendo uso del TP, se puede ver que el 25 % de los usuarios, es decir, dos de ellos recibieron una calificación de **bien** o 3 puntos según la escala, cuatro de ellos un 50 % de la población obtuvieron calificación de **muy bien**, 4 puntos según la escala y los últimos dos, el 25 % una calificación de **excelente**, 5 puntos en la escala. En comparación con la gráfica de la fig. 3.4 en la cual se muestran los resultados de la precisión en el *pick* para agarre correcto del objeto haciendo uso de la interfaz de voz, que cuatro de los usuarios un 50 % obtuvieron una calificación de **muy bien** y los otros cuatro el 50 % de la población obtuvieron una calificación de **excelente**. Con esto, se puede evidenciar que agarrando correctamente el objeto los usuarios recibieron mejores apreciaciones haciendo uso de la interfaz de voz.

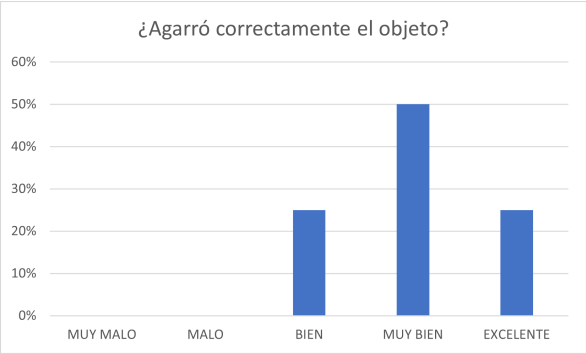


Figura 3.3: Resumen de resultados de la variable cualitativa precisión en el pick en el agarre correcto del objeto de los 8 usuarios usando el Teaching Pendant

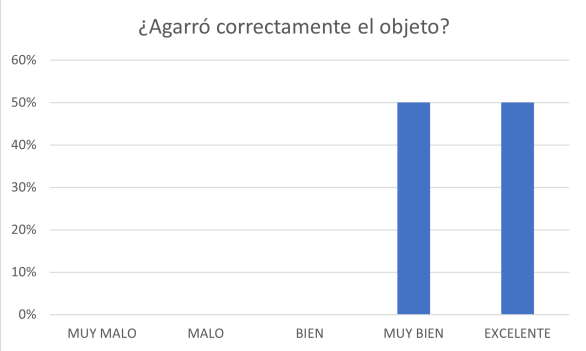


Figura 3.4: Resumen de resultados de la variable cualitativa precisión en el pick en el agarre correcto del objeto de los 8 usuarios usando la interfaz de voz

Se puede ver en la gráfica de barras de la fig. 3.5 los resultados de la precisión en el *pick* recogiendo el objeto correcto haciendo uso del TP, se puede ver que el 62,5 % de los usuarios, es decir, para cinco de ellos se calificó **muy bien** o 4 puntos según la escala y los restantes 3 el 37,5 % se calificó **excelente** 5 puntos en la escala. En comparación con la gráfica de la fig. 3.6 en la cual se muestran los resultados de la precisión en el *pick* recogiendo el objeto correcto haciendo uso de la interfaz de voz, que 2 de los usuarios un 25 % obtuvieron una calificación de **muy bien** y los otros 6 el 75 % de la población obtuvieron una calificación de **excelente**. Con esto, se puede evidenciar que en este aspecto los usuarios recibieron mejores apreciaciones haciendo uso de la interfaz de voz, fue más preciso el robot recogiendo la ficha correcta cuando se usa la interfaz de voz para esta tarea específica.

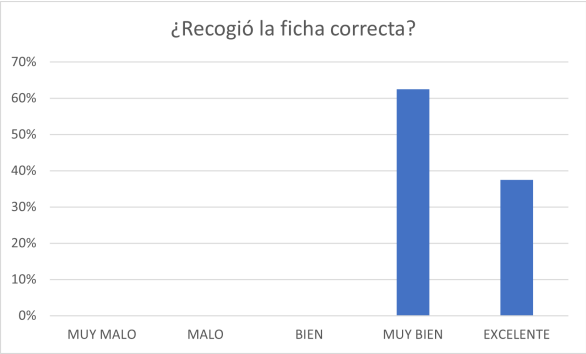


Figura 3.5: Resumen de resultados de la variable cualitativa precisión en el pick en la recogida de la ficha correcta de los 8 usuarios usando el Teaching Pendant

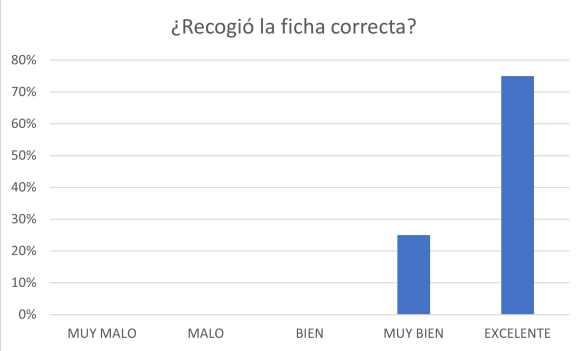


Figura 3.6: Resumen de resultados de la variable cualitativa precisión en el pick en la recogida de la ficha correcta de los 8 usuarios usando la interfaz de voz

En la gráfica de barras de la fig. 3.7 se muestran los resultados de la precisión en el *place* de la distancia del punto de referencia a donde fue colocada la ficha haciendo uso del TP, se puede ver que el 12,5 % de los usuarios, es decir, uno de ellos recibieron una calificación de **bien** o 3 puntos según la escala, seis de los usuarios es decir un 75 % de la población obtuvieron una calificación de **muy bien** 4 puntos de la escala y uno el 12,5 % de la población recibió una calificación de **excelente** 5 puntos en la escala. En comparación con la gráfica de la fig. 3.8 en la cual se muestran los resultados de la precisión en el *place* de la distancia del punto de referencia a donde fue colocada la ficha haciendo uso de la interfaz de voz, que siete de los usuarios un 87,5 % obtuvieron una calificación de **muy bien** y uno el 12,5 % de la población obtuvieron una calificación de **excelente**. En resumen, se obtuvo mejores resultados con la interfaz de voz en esta evaluación ya que con el TP se obtuvo una calificación de **bien** y con la interfaz de voz ningún usuario recibió esta calificación.

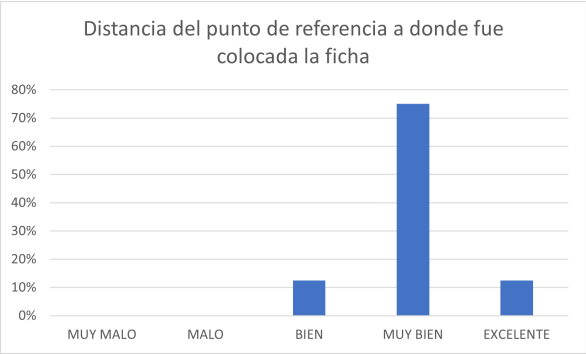


Figura 3.7: Resumen de resultados de la variable cualitativa precisión en el place en la distancia al punto de referencia de los 8 usuarios usando el Teaching Pendant

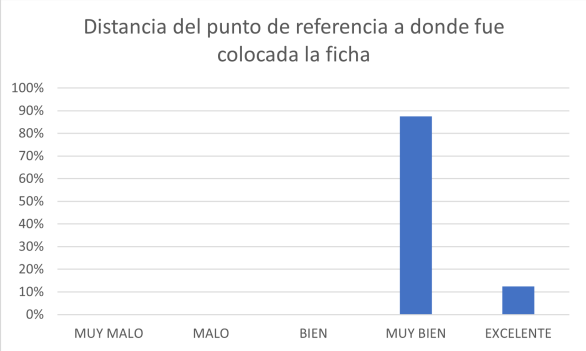


Figura 3.8: Resumen de resultados de la variable cualitativa precisión en el place en la distancia al punto de referencia de los 8 usuarios usando la interfaz de voz

En la gráfica de barras de la fig. 3.9 se muestran los resultados de la precisión en el *place* cuando el gripper suelta la ficha haciendo uso del TP, se puede ver que el 12,5 % de los usuarios, es decir, 1 de ellos recibieron una calificación de **bien** o 3 puntos según la escala, 2 de los usuarios es decir un 25 % de la población obtuvieron una calificación de **muy bien** 4 puntos de la escala y 5 el 62,5 % de la población recibió una calificación de **excelente** 5 puntos en la escala. En comparación con la gráfica de la fig. 3.10 en la cual se muestran los resultados de la precisión en el *place* cuando el gripper suelta la ficha haciendo uso de la interfaz de voz, que 1 de los usuarios un 12,5 % obtuvieron una calificación de **muy bien** y 7 el 87,5 % de la población obtuvieron una calificación de **excelente**. Por consiguiente, se obtuvo mejores resultados con la interfaz de voz en esta evaluación ya que con el TP se obtuvo una calificación de **bien** y con la interfaz de voz ningún usuario recibió esta calificación y la mayoría de calificaciones 7 de ellas fueren **excelente** en la interfaz de voz.

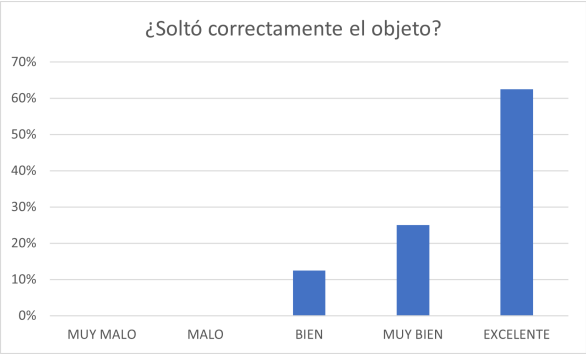


Figura 3.9: Resumen de resultados de la variable cualitativa precisión en el place en el desagarre correcto del objeto de los 8 usuarios usando el Teaching Pendant

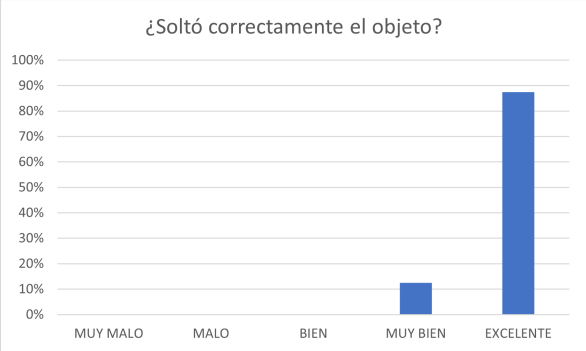


Figura 3.10: Resumen de resultados de la variable cualitativa precisión en el place en el desagarre correcto del objeto de los 8 usuarios usando la interfaz de voz

Se muestran en la gráfica de barras de la fig. 3.11 los resultados de la precisión en el *place* cuando se deja la ficha en la posición correcta haciendo uso del TP, se puede ver que el 37,5 % de los usuarios, es decir, 3 de ellos recibieron una calificación de **bien** o 3 puntos según la escala, 4 de los usuarios es decir un 50 % de la población obtuvieron una calificación de **muy bien** 4 puntos de la escala y 1 el 12,5 % de la población recibió una calificación de **excelente** 5 puntos en la escala. En comparación con la gráfica de la fig. 3.12 en la cual se muestran los resultados de la precisión en el *place* cuando se deja la ficha en la posición correcta haciendo uso de la interfaz de voz, que 6 de los usuarios un 75 % obtuvieron una calificación de **muy bien** y 2 el 25 % de la población obtuvieron una calificación de **excelente**. Por consiguiente, se obtuvo mejores resultados con la interfaz de voz en esta evaluación ya que con el TP se obtuvieron calificaciones de **bien** y con la interfaz de voz ningún usuario recibió esta calificación y se obtuvieron más calificaciones de **excelente** con la interfaz de voz.

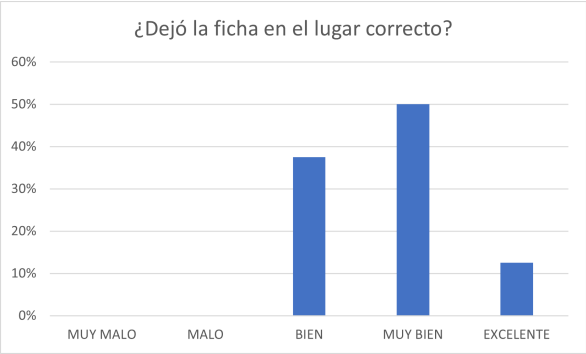


Figura 3.11: Resumen de resultados de la variable cualitativa precisión en el place en dejar la ficha en la posición correcta de los 8 usuarios usando el Teaching Pendant

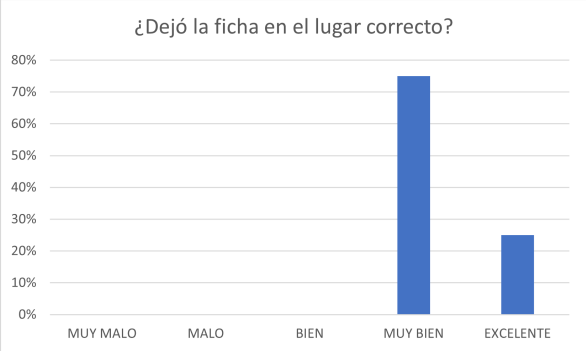


Figura 3.12: Resumen de resultados de la variable cualitativa precisión en el place en dejar la ficha en la posición correcta de los 8 usuarios usando la interfaz de voz

Como se evidencia en la tabla 3.5 todos los usuarios eligieron la interfaz de voz por encima de la programación habitual del TP que tiene el robot, para los usuarios con los que se hicieron las pruebas la interfaz de voz es mucho mas cómoda para desarrollar tareas que el TP. Dando la certeza que para este tipo de pruebas y este tipo de usuarios se obtuvo una elección del 100 % en la comodidad para la interfaz de voz, se espera que si se tratase de usuarios con diferentes características y condiciones la interfaz de voz también sea elegida en comodidad por encima del TP. Este resultado se esperaba en poca medida ya que para los humanos es mucho más natural hablarle a un robot que controlarlo por comandos de lenguaje gráfico. Inclusive, viviendo el auge de la utilización en gran masa en el primer mundo de los asistentes virtuales el hecho de comandar por voz ya es una realidad que debemos iniciar a recibir en mayor medida.

¿Cuál de las dos opciones te entregó mayor comodidad para desarrollar la actividad?		
Usuarios	**Prueba haciendo uso del TP**	**Prueba haciendo uso de la interfaz de voz**
Usuario 1		X
Usuario 2		X
Usuario 3		X
Usuario 4		X
Usuario 5		X
Usuario 6		X
Usuario 7		X
Usuario 8		X
Porcentaje de elección	0 %	100 %

Cuadro 3.5: Tabla con el resumen de la elección de los usuarios al preguntarles acerca de con cual de las dos interfaces se sentía más cómodo.

Como se muestra en la tabla 3.6 la mayoría de los usuarios (7 de los 8) eligieron a la interfaz de voz por encima del TP en la cualidad de interactividad con el robot UR3. Este resultado muestra que si es posible tener una mejor interacción con el robot gracias a esta implementación de la interfaz de voz, hay que trabajar en que la elección en esta cualidad sea unánime y que la interfaz de voz ayude más y más a tener una mejor interacción con el robot pensando en que los usuarios del robot sean muchos más y no un grupo corto de estudiantes de ingeniería en pregrado y postgrado.

¿Cuál de las dos opciones te permitió interactuar de la mejor forma con el robot UR3?		
Usuarios	**Prueba haciendo uso del TP**	**Prueba haciendo uso de la interfaz de voz**
Usuario 1		X
Usuario 2		X
Usuario 3		X
Usuario 4		X
Usuario 5		X
Usuario 6		X
Usuario 7	X	
Usuario 8		X
Porcentaje de elección	12.5 %	87.5 %

Cuadro 3.6: Tabla con el resumen de la elección de los usuarios al preguntarles acerca de que cual de las dos interfaces permite una mejor interacción con el robot.

Por parte del control comparando las dos interfaces la gráficas y la de voz en la tabla 3.7, se

evidencia que la elección por parte de los 8 usuarios fue equitativa, 4 eligieron el TP por ende se les facilitó el control con el Teaching Pendant. Por el contrario, a 4 de los 8 se les facilitó el control por medio de la interfaz de voz, esto denota es una alerta para la mejora en el manejo de la interfaz de voz. Algunos usuarios manifestaban que el robot no se movían cuando ellos les daban la orden, en algunos casos los usuarios decían sus comandos de voz cuando ya el programa había cerrado el micrófono que servía para la adquisición de la señal de voz, en otros casos los usuarios hacían pausas en el dictado de sus comandos de voz y el programa está diseñado para detectar esos espacios después del habla como finales de los comandos de voz, como el punto y aparte al finalizar un párrafo. También, la baja potencia en la voz, la posición al hablar, algunos le hablaban al computador teniendo el computador a su lado opuesto al que dirigían su voz. Estos fueron algunos fallos por parte de los usuarios todos estos errores fueron corregidos durante la prueba por falta de tiempo, pero se lograron disminuir en gran medida. Todas estas recomendaciones para el buen uso de la interfaz se compartieron con el usuario antes de la prueba, pero estos errores se asocian a que fue la primera vez que se encontraban con la interfaz de voz y si hubiesen contado con más tiempo algunos de los usuarios, con hacer las 3 pruebas se volvían expertos en el manejo del robot con la interfaz de voz, como lo hizo la usuaria 1 que realizó las 3 pruebas.

¿Cuál de las dos opciones te entregó mayor facilidad de control para llevar a cabo la actividad?		
Usuarios	**Prueba haciendo uso del TP**	**Prueba haciendo uso de la interfaz de voz**
Usuario 1		X
Usuario 2	X	
Usuario 3		X
Usuario 4	X	
Usuario 5		X
Usuario 6	X	
Usuario 7	X	
Usuario 8		X
Porcentaje de elección	50 %	50 %

Cuadro 3.7: Tabla con el resumen de la elección de los usuarios al preguntarles acerca de con cual de las dos interfaces se tenia mayor facilidad de control para el desarrollo de la actividad.

Para la pregunta: ¿Con cuál de las dos opciones crees que te demorarías menos en realizar una actividad con un grado de dificultad mayor?, en la tabla 3.8 se obtuvo una elección para la interfaz de voz del 87,5 %, 7 de los 8 usuarios decidieron que para realizar una actividad que demande mayor dificultad elegiría la interfaz de voz y una persona el 12,5 % eligió que con el Teaching Pendant se demoraría menos en realizar una actividad que demande más dificultad. Haciendo una perspectiva y mirando los resultados de las tablas cuantitativas de los tiempos 3.1 y 3.2 sabemos que con la interfaz de voz la prueba 1 que fue la única documentad en estos resultados ya que se podían comparar con

todos los usuarios, se dice que con la interfaz de voz se demoraron menos los 8 usuarios, por ende se espera que para una tarea que demande mayor dificultad la interfaz de voz sea la mejor, pero en el momento de la encuesta todas las respuestas fueron respetadas y en conclusión la mayoría eligió a la interfaz de voz. Entonces, se dice que es mejor la interfaz de voz para el desarrollo de esta actividad en cuestión.

¿Con cuál de las dos opciones crees que te demorarías menos en realizar una actividad con un grado de dificultad mayor?		
Usuarios	**Prueba haciendo uso del TP**	**Prueba haciendo uso de la interfaz de voz**
Usuario 1		X
Usuario 2		X
Usuario 3		X
Usuario 4		X
Usuario 5	X	
Usuario 6		X
Usuario 7		X
Usuario 8		X
Porcentaje de elección	12.5 %	87.5 %

Cuadro 3.8: Tabla con el resumen de la elección de los usuarios al preguntarles acerca de que con cual de las dos interfaces se demorarían menos realizando una actividad de mayor dificultad.

Cabe resaltar que las pruebas de la interfaz de voz implementada en el robot UR3 de la universidad se pensó en un inicio hacerlas con un público objetivo diferente al cual se realizaron las pruebas. Aún así, se puede concluir que se obtuvo una realimentación positiva de la interfaz de voz, tres de los ocho usuarios eligieron la interfaz de voz con una elección del 100 %, 4 de los 8 eligieron la interfaz con una elección del 75 % y 1 de los 8 usuarios decidieron calificar equivalentemente a las dos interfaces. Por tanto, se cumple con la hipótesis que planteaba una mejora en el manejo del robot en términos cuantitativos y cualitativos, cuantitativamente se puede ver una mejora en la reducción del tiempo para el desarrollo de una tarea especifica, en el número de intentos se logra dimensionar que ambas interfaces tiene la misma efectividad en la cantidad de intentos para la realización de cada movimiento y en el número de instrucciones se logra una mejora con la interfaz de voz implementada. Por parte, de las características cualitativas como lo es la precisión en el *pick* y en el *place*, se logra una mejora notable con la interfaz de voz, estas mejoras se documentan específicamente para estos usuarios, para este tipo de pruebas condicionadas a las limitaciones y adecuaciones que presentó esta implementación, puede que para otro tipo de implementación, pruebas o usuarios se presenten diferentes resultados pero se espera que sean en el mismo sentido que los resultados presentados en este documento.

Conclusiones

En este proyecto de grado se desarrolló una interfaz de voz capaz de recibir diferentes comandos de voz definidos en un diccionario de comandos seccionado en tres tipos de comandos de los cuales se puede componer una frase que seria el comando de voz asociado a un movimiento del robot UR3, con estos comandos el robot realiza tareas básicas de agarre, manipulación y desplazamiento de objetos, específicamente cubos hechos en el CAP en impresión 3d en material PLA. Este desarrollo contó con cuatros etapas principales: **Identificación**, **Definición**, **Implementación** y **Evaluación**, las cuales se detallan a continuación.

En este proyecto de grado se identificaron las principales interfaces de voz que se han implementado en brazos roboticos y publicado en diferentes revistas relevantes para la investigación, con esto se definieron las características que debía tener la interfaz de voz, como lo fue un buen dispositivo de adquisición de señales de audio, con reducción de ruido, buena ganancia en la recepción y un buen preprocesamiento de la señal. También, se definió la plataforma o lenguaje de programación a usar, se pensó en usar Alexa pero para esto se debía comprar un asistente virtual y todos los paquetes en la nube (Amazon Web Sevice AWS) necesarios para obtener la robustez deseada para la interfaz. Microsoft Speech Api 11 ofrece un reconocimiento de voz y una generación de voz, pero al ser una plataforma se debía instalar y adecuar un código para que todas las otras etapas de la interfaz de voz (movimientos del robot, decisiones, etc) pudiesen funcionar. Por ultimo, se planteó usar ROS y tener ventajas en el manejo del robot en diferentes lenguajes como lo hicieron Hofer y Strohmeier, pero ROS al ser un sistema operativo de robots tan potente y amplio, se debían modificar las tareas ya planteadas afectando así los tiempos de entrega ya afectados por situaciones externas al planteamiento de las tareas ya definidas en el anteproyecto. Por tal razón se eligió realizar la interfaz de voz en el lenguaje de programación python, este provee librerías que ofrecen grandes ventajas para el reconocimiento y generación de voz, dejando que el trabajo central fuese la comunicación y envió de información al robot, como lo son los movimientos y el manejo del gripper que fue lo que más tomó tiempo.

En este proyecto de grado se definió un diccionario de comandos con tres tipos de comandos, unos de movimiento (verbos de acción), otros de dirección y los de detalles, los cuales unidos forman las diferentes frases que ayudarían a la versatilidad de los comandos de voz y por ende la versatilidad del lenguaje de la interfaz de voz. Se definieron diferentes palabras para una misma acción para tener una banco de palabras mayor y poder armar la frase con diferentes variaciones que comprendería el comando de voz, así el usuario no diría la misma frase siempre, ya que naturalmente no repetimos la misma frase dos veces continuas.

En este trabajo de grado se implementó una interfaz de voz en el robot UR3 del CAP, la cual con ayuda de comandos de voz ya definidos, procesa la señal de audio adquirida y toma decisiones precisas que llevan ya sea al movimiento del robot o al movimiento de la pinza o gripper. Esta

implementación ayuda a dotar el robot de nuevas características que son naturalmente ajenas a el, las cuales son la interacción auditiva con un humano, ayudando a que personas que por algún motivo no puedan programar o manejar el robot gráficamente, lo puedan hacer por medio de su voz, aumentando así el publico objetivo de este robot.

En este trabajo de grado se evaluó la interfaz de voz implementada en el robot UR3 del CAP, elaborando un protocolo de pruebas en el cual se midieron características cuantitativas y cualitativas, y se comparó la interfaz de voz con la interfaz gráfica(TP), la cual provee el robot. Cuantitativamente la interfaz de voz presenta algunas ventajas frente a su contraparte en algunos casos, y cualitativamente se presenta la misma conclusión. Así como se presentan usuarios que mejoran su manejo del robot con la interfaz gráfica dedicando más tiempo en aprender su lenguaje de programación, así mismo se puede presentar que un usuario que maneje más de una vez la interfaz de voz, maneje mucho mejor el robot que su primera vez. Así, fue el caso de la usuaria 1, la cual fue la única que realizó las 3 pruebas propuestas con el robot y se puede ver mejoras en el manejo con el robot, pensando en que cada prueba iba aumentando su dificultad y tiempo de realización de la prueba.

Trabajos futuros

Los trabajos futuros irían encaminados a potenciar el reconocimiento del habla y a los sistemas de diálogos que puede incluir la interfaz para mejorar su interacción con al usuario. Se puede generar un nuevo modelo de dialogo, integrando técnicas de procesamiento del lenguaje natural, como también una integración de librerías en el programa que faciliten el planteamiento de nuevos comando de voz, formando oraciones con variaciones en los sinónimos que se pueden usar para la conformación de cada comando de voz.

Pensando en la integración de nuevas técnicas de dialogo también se puede trabajar con técnicas que traten multimodales, potenciando mucho más la interfaz de voz, pudiendo mezclar señales de voz con diferentes señales como imágenes, vídeo o incluso gestos. Así, se le puede dotar al robot de una versatilidad de manejo muy importante, ya que cualquier tipo de usuario podría manejar el robot, aumentando mucho más el publico objetivo del UR3 del CAP.

Por último, un trabajo mucho más amplio y grande se puede realizar aplicando una arquitectura cognitiva en el robot para que la resolución de diálogos sea mucho más orientada a un tipo de dialogo con un humano, esto se daría implementando inteligencia artificial y ciencia cognitiva computacional, el principal objetivo de este futuro trabajo es dotar al robot de una "mente" para que este pueda realizar un aprendizaje activo de diferentes tipos de tareas solo con la repetición de estas. Con esto, al momento de interactuar con el robot no tendríamos que repetir cada vez los mismos comandos cuando se realice una tarea que sea cotidiana o repetitiva en el tiempo para el robot.

Anexos

Como parte de anexos se tiene el código de la interfaz de voz que anteriormente ya se presentó en la etapa 2.3.2 de implementación ahí se encuentra el link del repositorio en GitHub pero se comparte nuevamente en esta capÍtulo, también se comparten los vídeos de las pruebas con los usuario y algunas pruebas hechas por mi. En algunas pruebas se colocó el micrófono de un lado por el espacio que tenia la mesa de trabajo por tal razón el usuario debía hablar de lado, de frente al micrófono, ya que la recomendación es hablar de frente al micrófono. También entre más cerca se hable al micrófono mejor es la adquisición de la señal de audio.

REPOSITORIO DEL CODIGO:

https://github.com/SebastianBlandon/UR3_VOICE_INTERFACE

REPOSITORIO DE LOS VIDEOS:

https://javerianacaliedu-my.sharepoint.com/:f:/g/personal/blandonsebas9715_javerianacali_edu_co/EkKqiq57R0hJpXDZKkgpglgBixYu4b3kW15L7URLQ3ZztA?e=O7i4sJ

Bibliografía

[1] S. Vicente. et al. Ejemplo de integración de alexa con un robot ur. 2019.

[2] Carl M. Rebman, Milam W. Aiken, and Casey G. Cegielski. Speech recognition in the human–computer interface. *Information & Management*, 40(6):509–519, July 2003.

[3] Abhinav Agrawal, D. Kyle Hogarth, and Septimiu Murgu. Robotic bronchoscopy for pulmonary lesions: a review of existing technologies and clinical data. *Journal of Thoracic Disease*, 12(6):3279–3286, June 2020.

[4] Mario Gaudino, Faisal Bakaeen, Piroze Davierwala, Antonino Di Franco, Stephen E. Fremes, Nirav Patel, John D. Puskas, Marc Ruel, Gianluca Torregrossa, Michael Vallely, and David P. Taggart and. New strategies for surgical myocardial revascularization. *Circulation*, 138(19):2160–2168, November 2018.

[5] Joshua A. Lee, Young Jae Byun, Shaun A. Nguyen, Eric J. Lentsch, and M. Boyd Gillespie. Transoral robotic surgery versus plasma ablation for tongue base reduction in obstructive sleep apnea: Meta-analysis. *Otolaryngology–Head and Neck Surgery*, 162(6):839–852, March 2020.

[6] M Diana and J Marescaux. Robotic surgery. *British Journal of Surgery*, 102(2):e15–e28, January 2015.

[7] Steven Smith. *Digital signal processing : a practical guide for engineers and scientists*. Newnes, Amsterdam Boston, 2003.

[8] Jinyu Li, Li Deng, Reinhold Haeb-Umbach, and Yifan Gong. Introduction. In *Robust Automatic Speech Recognition*, pages 1–7. Elsevier, 2016.

[9] Alexandros Tsilfidis, Iosif Mporas, John Mourjopoulos, and Nikos Fakotakis. Automatic speech recognition performance in different room acoustic environments with and without dereverberation preprocessing. *Computer Speech & Language*, 27(1):380–395, January 2013.

[10] A. Waibel, P. Geutner, L.M. Tomokiyo, T. Schultz, and M. Woszczyna. Multilinguality in speech and spoken language systems. *Proceedings of the IEEE*, 88(8):1297–1313, August 2000.

[11] S. Wermter and V. Weber. Interactive spoken-language processing in a hybrid connectionist system. *Computer*, 29(7):65–74, July 1996.

[12] R. Cole, L. Hirschman, L. Atlas, M. Beckman, A. Biermann, M. Bush, M. Clements, L. Cohen, O. Garcia, B. Hanson, H. Hermansky, S. Levinson, K. McKeown, N. Morgan, D.G. Novick, M. Ostendorf, S. Oviatt, P. Price, H. Silverman, J. Spiitz, A. Waibel, C. Weinstein, S. Zahorian, and V. Zue. The challenge of spoken language systems: Research directions for the nineties. *IEEE Transactions on Speech and Audio Processing*, 3(1):1–21, 1995.

[13] Sangkeun Jung, Cheongjae Lee, Seokhwan Kim, and Gary Geunbae Lee. DialogStudio: A workbench for data-driven spoken dialog system development and management. *Speech Communication*, 50(8-9):697–715, August 2008.

[14] Victor Zue and Stephanie Seneff. Spoken dialogue systems. In *Springer Handbook of Speech Processing*, pages 705–722. Springer Berlin Heidelberg, 2008.

[15] David Griol, Zoraida Callejas, Ramón López-Cózar, and Giuseppe Riccardi. A domain-independent statistical methodology for dialog management in spoken dialog systems. *Computer Speech & Language*, 28(3):743–768, May 2014.

[16] Douglas O'Shaughnessy. Invited paper: Automatic speech recognition: History, methods and challenges. *Pattern Recognition*, 41(10):2965–2979, October 2008.

[17] Ramón López-Cózar and Zoraida Callejas. ASR post-correction for spoken dialogue systems based on semantic, syntactic, lexical and contextual information. *Speech Communication*, 50(8-9):745–766, August 2008.

[18] Wei-Lin Wu, Ru-Zhan Lu, Jian-Yong Duan, Hui Liu, Feng Gao, and Yu-Quan Chen. Spoken language understanding using weakly supervised learning. *Computer Speech & Language*, 24(2):358–382, April 2010.

[19] Ramón López-Cózar, Zoraida Callejas, and David Griol. Using knowledge of misunderstandings to increase the robustness of spoken dialogue systems. *Knowledge-Based Systems*, 23(5):471–485, July 2010.

[20] David R. Traum and Staffan Larsson. The information state approach to dialogue management. In *Text, Speech and Language Technology*, pages 325–353. Springer Netherlands, 2003.

[21] Jason D. Williams and Steve Young. Partially observable markov decision processes for spoken dialog systems. *Computer Speech & Language*, 21(2):393–422, April 2007.

[22] David Griol, Lluís F. Hurtado, Encarna Segarra, and Emilio Sanchis. A statistical approach to spoken dialog systems design and evaluation. *Speech Communication*, 50(8-9):666–682, August 2008.

[23] Oliver Lemon. Learning what to say and how to say it: Joint optimisation of spoken dialogue management and natural language generation. *Computer Speech & Language*, 25(2):210–221, April 2011.

[24] Víctor López Salazar, Eduardo M. Eisman Cabeza, Juan Luis Castro Peña, and Jose Manuel Zurita López. A case based reasoning model for multilingual language generation in dialogues. *Expert Systems with Applications*, 39(8):7330–7337, June 2012.

[25] Thierry Dutoit. *An introduction to text-to-speech synthesis*, volume 3. Springer Science & Business Media, 1997.

[26] Jinyu Li, Li Deng, Reinhold Haeb-Umbach, and Yifan Gong. Fundamentals of speech recognition. In *Robust Automatic Speech Recognition*, pages 9–40. Elsevier, 2016.

[27] Valeria Villani, Fabio Pini, Francesco Leali, Cristian Secchi, and Cesare Fantuzzi. Survey on human-robot interaction for robot programming in industrial applications. *IFAC-PapersOnLine*, 51(11):66–71, 2018.

[28] David Hinwood, James Ireland, Elizabeth Ann Jochum, and Damith Herath. A proposed wizard of OZ architecture for a human-robot collaborative drawing task. In *Social Robotics*, pages 35–44. Springer International Publishing, 2018.

[29] Damith C. Herath, Elizabeth Jochum, and Evgenios Vlachos. An experimental study of embodied interaction and human perception of social presence for interactive robots in public settings. *IEEE Transactions on Cognitive and Developmental Systems*, 10(4):1096–1105, December 2018.

[30] Jan Czarnowski, Adam Dąbrowski, Mateusz Maciaś, Jakub Główka, and Józef Wrona. Technology gaps in human-machine interfaces for autonomous construction robots. *Automation in Construction*, 94:179–190, October 2018.

[31] Daeho Kim, Ankit Goyal, Alejandro Newell, SangHyun Lee, Jia Deng, and Vineet R. Kamat. Semantic relation detection between construction entities to support safe human-robot collaboration in construction. In *Computing in Civil Engineering 2019*. American Society of Civil Engineers, June 2019.

[32] Patrik Gustavsson, Magnus Holm, Anna Syberfeldt, and Lihui Wang. Human-robot collaboration – towards new metrics for selection of communication technologies. *Procedia CIRP*, 72:123–128, 2018.

[33] Julia Berg, Albrecht Lottermoser, Christoph Richter, and Gunther Reinhart. Human-robot-interaction for mobile industrial robot teams. *Procedia CIRP*, 79:614–619, 2019.

[34] M. Beaudouin-Lafon. An overview of human-computer interaction. *Biochimie*, 75(5):321–329, January 1993.

[35] G. Fraser Shein, Jutta Treviranus, Nicholas D. Brownlow, Morris Milner, and Penny Parnes. An overview of human-computer interaction techniques for people with physical disabilities. *International Journal of Industrial Ergonomics*, 9(2):171–181, February 1992.

[36] Philip Tucker and Dylan M. Jones. Voice as interface: An overview. *International Journal of Human-Computer Interaction*, 3(2):145–170, January 1991.

[37] Michael H Cohen, Michael Harris Cohen, James P Giangola, and Jennifer Balogh. *Voice user interface design*. Addison-Wesley Professional, 2004.

[38] P. Biel. et al. *A tour-guide robot: Moving towards interaction with humans*. 2020.

[39] Veton Kepuska and Gamal Bohouta. Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home). In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, January 2018.

[40] F. Åsa. et al. *Evaluating Cobots For Final Assembly*. 2016.

[41] C. Araceli. *Control de robot Mitsubishi (RV M1) por medio de una interfaz verbal*. 2004.

[42] P. Mario. *Reconocimiento de voz con el robot Félix*. 2016.

[43] S. Cecilia. et al. *Control de una silla robótica a través de comandos de voz*. 2012.

[44] W. Naoki. et al. *Enhancing Listening Capability of Humanoid Robot by Reduction of Stationary Ego-Noise*. 2019.

[45] F. Hofer, D. Strohmeier. *Multilingual speech control for ROS-driven robots*. 2019.

[46] G. Patrik. et al. *Human-Robot Collaboration Demonstrator Combining Speech Recognition and Haptic Control*. 2017.

[47] Valerio Cornagliotto, Elisa Digo, and Stefano Pastorelli. Using a robot calibration approach toward fitting a human arm model. In *Advances in Service and Industrial Robotics*, pages 199–207. Springer International Publishing, 2021.

[48] Kunxia Wang, Guoxin Su, Li Liu, and Shu Wang. Wavelet packet analysis for speaker-independent emotion recognition. *Neurocomputing*, 398:257–264, July 2020.

[49] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, March 1998.

[50] Biswajit Karan, Kartik Mahto, and Sitanshu Sekhar Sahu. Intelligent speech processing in the time-frequency domain. In *Intelligent Speech Signal Processing*, pages 153–173. Elsevier, 2019.

[51] Konstantin Dragomiretskiy and Dominique Zosso. Variational mode decomposition. *IEEE Transactions on Signal Processing*, 62(3):531–544, February 2014.

[52] AdvanceTEC Cleanroom Integration. https://www.advancetecllc.com/classifications/iso-5-cleanroom-classification, 22 de Junio de 2021.

[53] ALLPE. https://www.allpe.com/acustica/ingenieria-acustica/mediciones-acusticas/a-que-equivalen-los-diferentes-niveles-de-decibelios/a-que-equivalen-65-decibelios/, 22 de Junio de 2021.

[54] Konrad Ahlin, Benjamin Joffe, Ai-Ping Hu, Gary McMurray, and Nader Sadegh. Autonomous leaf picking using deep learning and visual-servoing. *IFAC-PapersOnLine*, 49(16):177–183, 2016.