Durante el desarrollo de esta práctica se utilizo la paquetería de pocketsphinx para poder utilizar el reconocimiento de voz que permite utilizar. Para dicho reconocimiento es necesario utiliza una lista de palabras, un modelo de lenguaje o una gramática con su respectivo diccionario. Para la realización de esta práctica se utilizará una gramática y un diccionario.

Una gramática hace referencia a un archivo con un respectivo formato formado con reglas que definen las frases que podrán ser reconocidas por medio de la voz. Estas frases están definidas por un conjunto de palabras descritas en un diccionario, siendo esta último un archivo que provee al sistema de una transformación de un conjunto de palabras a una secuencia de fonemas.

Inicialmente, se ejecutó el comando *roslaunch bring up pocketsphinx_test.launch* con el cual se comienza el reconocimiento de voz. Este archivo fue modificado para seguir la gramática del archivo "restaurant.gram". Anteriormente, tenía una gramática de movimiento con comandos como: Move Rigth, Move Left, Move Foward Move Back, Full Speed, Half Speed.

Estas frases están definidas dentro de la siguiente gramática y su respectivo diccionario:

```
#JSGF V1.0;

/**
 * JSGF Grammar similar to kws example
 */

grammar voice_cmd;

public <move> = MOVE FORWARD;

public <move2> = MOVE <direction> [<speed>];

<direction> = FORWARD | BACK | LEFT | RIGHT;

<speed> = FULL SPEED | HALF SPEED;

```

```
BACK    B AE K
FORWARD F AO R W ER D
FULL    F UH L
HALF    HH AE F
LEFT    L EH F T
MOVE    M UW V
RIGHT   R AY T
SPEED   S P IY D
STOP    S T AA P
```

Del lado izquierdo se encuentra el archivo gram donde se definen las frases de la parte superior y del lado derecho se encuentran las palabras con sus respectivos fonemas del archivo dict.

Ejemplos de frases que se pueden reconocer de acuerdo con la gramática "restaurant.gram"

Posteriormente se cambiarían los parámetros para utilizar la gramática básica de un restaurante con la cual se podrían reconocer las frases típicas para ordenar alimento. La gramática definida en el archivo restaurant.gram es la siguiente:

```
grammar restaurant;
public <command> = (<justina> | <wantCombo> | <
wantBeverage>);
<justina> = JUSTINA (YES | NO | WAIT | START | TAKE A ORDER
| THIS IS THE ORDER);
<wantCombo> = I WANT [A] (PRINGLES | CHIPS | PASTA |
COOKIES | PASTA | NOODLES | TUNA FISH | PICKLES | CHOCO
FLAKES | ROBO O'S | MUESLI | M AND M) AND (PRINGLES | CHIPS
| PASTA | COOKIES | PASTA | NOODLES | TUNA FISH | PICKLES |
CHOCO FLAKES | ROBO O'S | MUESLI | M AND M);
<wantBeverage> = I WANT A (TEA | BEER | COKE | WATER);

```

Según la información de dicho archivo se podrían reconocer frases como las siguientes:

- Justina no
- Justina take a order
- Justina this is the order
- I want pringles and chips
-
-
- I want noodles and pasta
- I want m and m and m and m
- I want choco flakes and robo o´s
- I want pickles and pringles
- I want a tea
- I want a coke

Explicación breve de lo que significan los cuatro parámetros mencionados en el punto 5 de la diapositiva 19.

Durante el desarrollo de la práctica se modificaron algunos parámetros dentro del archivo *pocketsphinx_test.launch,* los cuales fueron los siguientes:

- ✓ *Cambiar el valor del parámetro gram de .../voice cmd a .../restaurant*

Al modificar este parámetro, se modifica el archivo donde se encuentra descrita la gramática de reconocimiento, es decir, se cambia de reconocer las frases de movimiento, a reconocer la frases del restaurante

- ✓ *Cambiar el valor del parámetro dict de ../voice cmd.dic a restaurant.dict*

De igual forma que el punto anterior, en esta línea se cambia el diccionario con palabras de movimiento a las del restaurante, es aquí donde se encuentran las palabras con sus respectivos fonemas.

- ✓ *Cambiar el valor de grammar de voice cmd a restaurant*

En esta línea, se identifica como tal el nombre de la gramática, este nombre es la primera línea dentro de los archivos gram, y es en este punto del archivo launch es donde identifica dicho nombre. A difencia de modificar la línea del parametro gram, aquí se cambia el nombre más no el archivo, podría haber dos gramáticas dentro del mismo archivo gram, pero se seleccionaría el nombre que se desee utilizar.

- ✓ *Cambiar el valor de rule de move2 a command*

En esta línea se seleccionan el conjunto de frases dentro de una gramática que se desea que sean reconocidas, este parámetro podría incluir 1 o más frases delimitas por pico paréntesis y líneas. En la imagen superior se observa como es la estructura de estos últimos dos puntos.

Capturas de pantalla donde se observen las frases reconocidas.

Las siguientes imagen muestra el resultado de que no se haya podido identificar alguna de las frases durante el reconocimiento, esto por una mala pronunciación o falta de alguna de las palabras dentro de una cierta frase restringida.

```
INFO: fsg_search.c(869): fsg 0.32 CPU 0.179 xRT
INFO: fsg_search.c(871): fsg 2.05 wall 1.167 xRT
ERROR: "fsg_search.c", line 940: Final result does not match the grammar i
n frame 175
```

Las siguientes son las salidas a diversas frases dentro del reconocimiento de voz.

```
INFO: fsg_search.c(869): fsg 0.18 CPU 0.186 xRT
INFO: fsg_search.c(871): fsg 2.63 wall 2.742 xRT
[INFO] [1586633970.009004]: OUTPUT: "I WANT A TEA"
INFO: cmn_live.c(88): Update from < 73.54 10.79 -4.44  6.49  2.99 10.82  3
.84  1.04 -0.52  4.12 -0.18 -1.98 -1.95 >
INFO: cmn_live.c(105): Update to   < 73.29 10.77 -4.27  6.35  3.02 10.81
3.87  1.05 -0.52  4.07 -0.10 -1.92 -1.90 >
INFO: cmn_live.c(120): Update from < 73.29 10.77 -4.27  6.35  3.02 10.81
3.87  1.05 -0.52  4.07 -0.10 -1.92 -1.90 >
INFO: cmn_live.c(138): Update to   < 74.65 10.65 -3.32  6.56  2.02 11.49
3.61  1.14 -0.96  2.66 -1.27 -2.39 -1.05 >
INFO: fsg_search.c(859): 175 frames, 15281 HMMs (87/fr), 33681 senones (19
2/fr), 3570 history entries (20/fr)
```

```
INFO: fsg_search.c(869): fsg 0.42 CPU 0.185 xRT
INFO: fsg_search.c(871): fsg 3.41 wall 1.495 xRT
[INFO] [1586633963.922140]: OUTPUT: "I WANT PASTA AND PICKLES"
INFO: cmn_live.c(88): Update from < 74.26 10.96 -3.73  4.80  2.67 10.77  4
.30  1.95 -0.19  6.01  3.32 -0.93 -3.19 >
INFO: cmn_live.c(105): Update to   < 74.49 11.28 -3.60  5.18  2.59 10.84
4.36  1.74 -0.06  6.09  3.01 -1.10 -3.14 >
INFO: cmn_live.c(120): Update from < 74.49 11.28 -3.60  5.18  2.59 10.84
4.36  1.74 -0.06  6.09  3.01 -1.10 -3.14 >
INFO: cmn_live.c(138): Update to   < 73.81 10.47 -4.22  6.91  2.46 10.60
3.84  1.24 -0.12  3.92 -0.38 -2.03 -1.81 >
INFO: fsg_search.c(859): 241 frames, 6844 HMMs (28/fr), 18205 senones (75/
fr), 1483 history entries (6/fr)
```

```
INFO: fsg_search.c(869): fsg 0.25 CPU 0.210 xRT
INFO: fsg_search.c(871): fsg 3.59 wall 3.070 xRT
[INFO] [1586633924.219308]: OUTPUT: "JUSTINA NO"
INFO: cmn_live.c(120): Update from < 67.81 11.27  3.38  9.79  4.16 10.31
7.14  5.75  3.89  2.77 -0.85  0.36 -1.58 >
INFO: cmn_live.c(138): Update to   < 72.65  8.96  0.03  8.14  4.01 11.18
4.01  2.98  2.02  3.27 -1.22 -1.61 -1.72 >
INFO: fsg_search.c(859): 222 frames, 15847 HMMs (71/fr), 32056 senones (14
4/fr), 3191 history entries (14/fr)
```

```
INFO: fsg_search.c(869): fsg 0.28 CPU 0.175 xRT
INFO: fsg_search.c(871): fsg 2.86 wall 1.785 xRT
[INFO] [1586633946.906498]: OUTPUT: "I WANT A COKE"
INFO: cmn_live.c(120): Update from < 73.40 11.07 -1.47  6.41  4.37 10.04
3.05  2.80  0.91  3.83 -1.37 -2.30 -1.35 >
INFO: cmn_live.c(138): Update to   < 73.60 11.32 -1.86  6.19  4.27 10.30
3.34  2.53  0.71  4.80 -0.28 -2.69 -2.46 >
INFO: fsg_search.c(859): 92 frames, 5172 HMMs (56/fr), 11695 senones (127/
fr), 1440 history entries (15/fr)
```

Cabe destacar que las frases con mayor número de palabras eran las más difíciles de detectar por el tiempo que da el sistema para decir la frase completa.