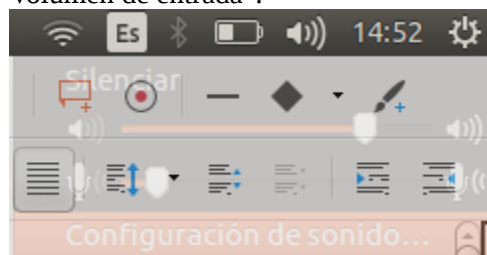


Reconocimiento de voz utilizando *pocketsphinx*

Para esta actividad primero configuramos la entrada de audio, ya sea por el micrófono interno de la computadora, o por auriculares que tienen incluido micrófono. En este caso de preferencia tomar la segunda opción para que el reconocimiento sea bastante cercano a lo que decimos realmente, al conectar los auriculares nos saldrá una nueva ventana emergente como la siguiente figura, y de preferencia seleccionamos la segunda opción.



Posteriormente nos vamos a la barra de estados, y seleccionamos el icono de volumen, le damos clic en "configuración de sonido...", ya estando dentro damos clic en "entrada", le damos clic en "Auriculares con micrófono" y luego calibramos en "Volumen de entrada".



Luego, nos dirigimos a la carpeta `catkin_ws/src/pocketsphinx/vocab/`, y abrimos el archivo `restaurant.gram`, y ahí revisamos la estructura gramatical que tiene el comando `<command>`, y de los elementos de la estructura las

palabras para formar frases posibles en el reconocimiento. Cabe destacar que en la gramática de “restaurant” no tiene una estructura fija, sino tiene tres estructuras gramaticales que puede reconocer.

```
public <command> = (<justina> | <wantCombo> | <wantBeverage>);
```

- <justina>: JUSTINA YES, JUSTINA TAKE A ORDER.
- <wantCombo>: I WANT COOKIES, I WANT PASTA AND PRINGLES.
- <wantBeverage>: I WANT A BEER, I WANT A COKE.

Anteriormente se mostraron de acuerdo a las estructuras gramaticales, dos ejemplos de las frases que se pueden formar de cada estructura. En la siguiente figura se muestra las posibles combinaciones de palabras que puede reconocer.

```
#JSGF V1.0;
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);
```

Ahora nos dirigimos a la carpeta catkin ws/src/bring up/launch/ y abrimos el archivo pocketsphinx_text.launch, y vamos a modificar los siguientes parámetros:

- Parámetro gram de ../voice_cmd a ../restaurant. El parámetro de gram significa el nombre del archivo en donde se estructura y establecen reglas de la gramática para el reconocimiento de voz.
- Parámetro dict de ../voice_cmd.dic a restaurant.dict. El parámetro de dict significa el nombre del archivo en donde se establece la fonética del vocabulario que es parte de la gramática.
- Parámetro grammar de voice_cmd a restaurant. El parámetro grammar quiere decir la confirmación del primer parámetro gram, ya que dentro del archivo indicado en el parámetro gram, se declara la gramática en este caso es por la línea de código dentro de restaurant: grammar restaurant; .
- Parámetro rule de move2 a command. Este parámetro es la selección, dentro del código de la gramática, de la estructura gramatical.

A continuación vamos a introducir la siguiente línea de código roslaunch bring_up pocketsphinx_test.launch, y una vez inicializado pocketsphinx, vamos a decir en el micrófono (de los manos libres) las frases que anteriormente se ejemplificaron.

```
INFO: fsg_search.c(869): fsg 0.10 CPU 0.070 xRT
INFO: fsg_search.c(871): fsg 2.95 wall 1.980 xRT
ERROR: "fsg_search.c", line 940: Final result does not match the grammar in frame 148
INFO: cmn_live.c(120): Update from < 56.46 2.93 1.10 16.35 -2.97 7.35 -7.54
4.39 -1.53 -2.90 1.99 -0.10 -0.58 >
INFO: cmn_live.c(138): Update to < 55.38 1.86 0.51 18.00 -3.59 6.54 -7.23
3.98 -1.23 -3.75 2.60 -0.44 -0.98 >
INFO: fsg_search.c(859): 152 frames, 1289 HMMs (8/fr), 5383 senones (35/fr), 336
history entries (2/fr)

INFO: fsg_search.c(869): fsg 0.08 CPU 0.055 xRT
INFO: fsg_search.c(871): fsg 3.00 wall 1.959 xRT
[INFO] [1586894092.098268]: OUTPUT: "JUSTINA YES"
```

```
INFO: cmn_live.c(88): Update from < 47.58 -1.35 -0.33 13.29 -3.50 4.79 -6.09 3
.43 -1.90 -2.43 2.68 0.43 -1.06 >
INFO: cmn_live.c(105): Update to < 49.38 0.51 -2.35 13.79 -5.01 5.23 -5.56
2.70 -1.39 -2.25 2.74 -0.92 -0.40 >
INFO: cmn_live.c(120): Update from < 49.38 0.51 -2.35 13.79 -5.01 5.23 -5.56
2.70 -1.39 -2.25 2.74 -0.92 -0.40 >
INFO: cmn_live.c(138): Update to < 49.61 2.54 -3.72 12.61 -5.37 5.73 -5.09
1.46 -1.07 -2.07 2.12 0.10 -0.08 >
INFO: fsg_search.c(859): 262 frames, 3930 HMMs (15/fr), 12866 senones (49/fr), 6
82 history entries (2/fr)

INFO: fsg_search.c(869): fsg 0.65 CPU 0.248 xRT
INFO: fsg_search.c(871): fsg 68.12 wall 25.901 xRT
[INFO] [1586894199.560171]: OUTPUT: "JUSTINA TAKE A ORDER"
```

```
INFO: cmn_live.c(120): Update from < 47.66 3.09 -3.69 10.98 -5.42 6.54 -3.93
1.22 -1.29 -1.70 2.43 -0.08 -1.10 >
INFO: cmn_live.c(138): Update to < 49.27 4.01 -3.43 9.33 -5.92 6.96 -4.41
2.05 -2.47 -2.12 3.56 -0.91 -1.20 >
INFO: fsg_search.c(859): 181 frames, 7304 HMMs (40/fr), 18970 senones (104/fr),
1511 history entries (8/fr)

INFO: fsg_search.c(869): fsg 0.16 CPU 0.088 xRT
INFO: fsg_search.c(871): fsg 3.37 wall 1.851 xRT
[INFO] [1586894209.939464]: OUTPUT: "I WANT COOKIES"
```

```
INFO: cmn_live.c(88): Update from < 47.60 9.19 -1.48 5.50 -7.22 8.04 -3.66 4
.02 -3.08 -1.32 3.03 -1.38 -1.33 >
INFO: cmn_live.c(105): Update to < 49.84 9.29 -2.67 5.62 -8.77 7.94 -3.84
4.23 -3.96 -0.18 3.39 -1.11 -1.74 >
INFO: cmn_live.c(120): Update from < 49.84 9.29 -2.67 5.62 -8.77 7.94 -3.84
4.23 -3.96 -0.18 3.39 -1.11 -1.74 >
INFO: cmn_live.c(138): Update to < 51.08 9.16 -2.62 5.05 -9.13 7.45 -2.88
3.37 -4.11 -0.46 3.38 -1.04 -1.70 >
INFO: fsg_search.c(859): 267 frames, 6294 HMMs (23/fr), 18080 senones (67/fr), 1
259 history entries (4/fr)

INFO: fsg_search.c(869): fsg 0.16 CPU 0.060 xRT
INFO: fsg_search.c(871): fsg 5.82 wall 2.171 xRT
[INFO] [1586894240.485413]: OUTPUT: "I WANT PASTA AND PRINGLES"
```

```
INFO: cmn_live.c(88): Update from < 51.08 9.16 -2.62 5.05 -9.13 7.45 -2.88 3
.37 -4.11 -0.46 3.38 -1.04 -1.70 >
INFO: cmn_live.c(105): Update to < 52.26 8.38 -4.27 7.09 -9.37 7.67 -4.21
2.58 -4.26 -0.42 3.14 -0.77 -1.90 >
INFO: cmn_live.c(120): Update from < 52.26 8.38 -4.27 7.09 -9.37 7.67 -4.21
2.58 -4.26 -0.42 3.14 -0.77 -1.90 >
INFO: cmn_live.c(138): Update to < 50.06 8.28 -4.71 7.13 -9.04 6.87 -3.64
1.95 -3.68 -0.17 2.10 -0.27 -2.09 >
INFO: fsg_search.c(859): 222 frames, 11176 HMMs (50/fr), 27567 senones (124/fr),
1616 history entries (7/fr)

INFO: fsg_search.c(869): fsg 0.22 CPU 0.097 xRT
INFO: fsg_search.c(871): fsg 4.95 wall 2.221 xRT
[INFO] [1586894245.439327]: OUTPUT: "I WANT A BEER"
```

```

INFO: cmn_live.c(88): Update from < 46.77  0.82 -4.04 11.68 -4.97  5.50 -5.10  1
.23 -1.01 -1.73  1.90  0.31 -0.09 >
INFO: cmn_live.c(105): Update to  < 49.65  2.43 -4.43 11.50 -6.52  6.41 -4.91
1.12 -1.71 -2.31  2.69 -0.53 -0.76 >
INFO: cmn_live.c(120): Update from < 49.65  2.43 -4.43 11.50 -6.52  6.41 -4.91
1.12 -1.71 -2.31  2.69 -0.53 -0.76 >
INFO: cmn_live.c(138): Update to  < 47.66  3.09 -3.69 10.98 -5.42  6.54 -3.93
1.22 -1.29 -1.70  2.43 -0.08 -1.10 >
INFO: fsg_search.c(859): 212 frames, 9455 HMMs (44/fr), 23622 senones (111/fr),
1945 history entries (9/fr)

INFO: fsg_search.c(869): fsg 0.15 CPU 0.069 xRT
INFO: fsg_search.c(871): fsg 3.83 wall 1.796 xRT
[INFO] [1586894206.569506]: OUTPUT: "I WANT A COKE"

```

```

INFO: cmn_live.c(88): Update from < 47.60  9.19 -1.48  5.50 -7.22  8.04 -3.66  4
.02 -3.08 -1.32  3.03 -1.38 -1.33 >
INFO: cmn_live.c(105): Update to  < 49.84  9.29 -2.67  5.62 -8.77  7.94 -3.84
4.23 -3.96 -0.18  3.39 -1.11 -1.74 >
INFO: cmn_live.c(120): Update from < 49.84  9.29 -2.67  5.62 -8.77  7.94 -3.84
4.23 -3.96 -0.18  3.39 -1.11 -1.74 >
INFO: cmn_live.c(138): Update to  < 51.08  9.16 -2.62  5.05 -9.13  7.45 -2.88
3.37 -4.11 -0.46  3.38 -1.04 -1.70 >
INFO: fsg_search.c(859): 267 frames, 6294 HMMs (23/fr), 18080 senones (67/fr), 1
259 history entries (4/fr)

INFO: fsg_search.c(869): fsg 0.16 CPU 0.060 xRT
INFO: fsg_search.c(871): fsg 5.82 wall 2.171 xRT
[INFO] [1586894240.485413]: OUTPUT: "I WANT PASTA AND PRINGLES"

```

Conclusiones

En la robótica y en la inteligencia artificial, el reconocimiento de voz es un enorme desafío para que el usuario se sienta cómodo al mandar ordenes de ejecución de un sistema. Es como una televisora, en vez de tener un teclado de control remoto, tendría un micrófono y en base a frases que yo pueda repetir como “el 5” o “canal 5”, “Nickelodeon” o “Nick”. Hoy en día los dispositivos electrónicos o maquinaria ya se esta implementando funcionamiento a través del conocimiento de voz.

En ésta practica vimos que ROS tiene una aplicación llamada pocketphnix que nos permite reconocer instrucciones dichas, pero esas instrucciones que reconoce, se establecen gracias a un archivo grammar y un archivo dictionary. Un archivo grammar establece la(s) estructura(s) de las frases que puede reconocer, y un elemento de esa estructura esta conformado por palabras. Un archivo dictionary establece la manera en como reconoce las palabras establecidas en el grammar.