



Interficies Multimodales basadas en Avatares Virtuales



Ingeniería Informática
Proyecto Fin de Carrera

Francisco Gutiérrez Romero

Barcelona, 19 de Septiembre 2011

ÍNDICE

❖ Introducción

- Avatar Virtual Proyecto Hermes

❖ Objetivos

❖ Estado del Arte

❖ Base del Proyecto

- XFace Toolkits Flite Julius Lenguaje Natural Base de datos

❖ Diseño e Implementación

- Entorno gráfico Text to Speech Speech to Text Interacción

❖ Resultados

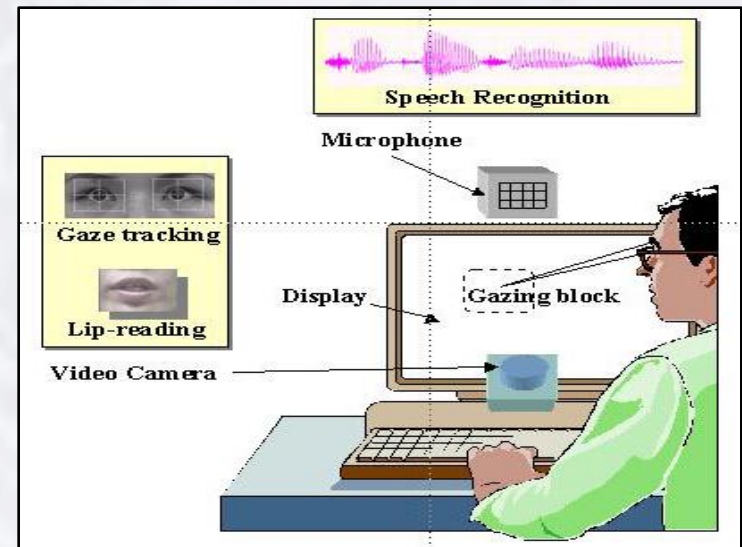
- Vídeo demostrativo

❖ Conclusión

- Conclusión Trabajo Futuro

INTRODUCCIÓN

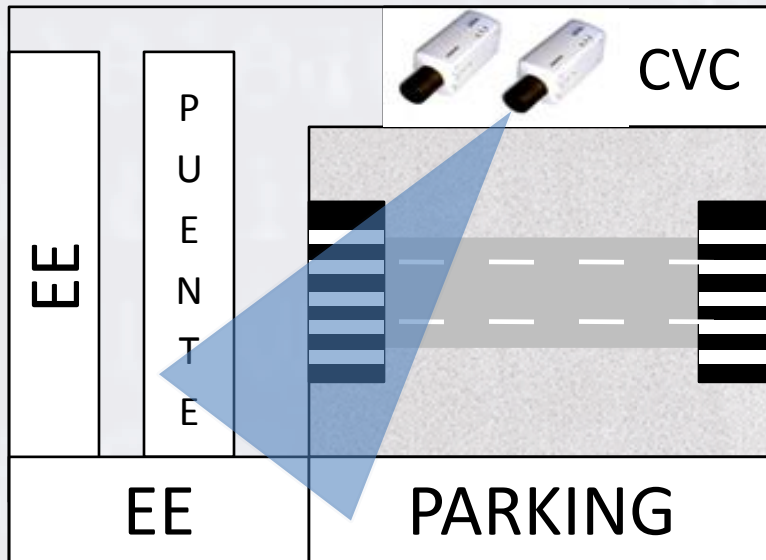
- Avatar Virtual: Representación gráfica asociada a un usuario.



- Multimodal: Comunicación en la que intervienen modos humanos, la voz o el habla, los gestos, los ojos, los movimientos, etc...

INTRODUCCIÓN

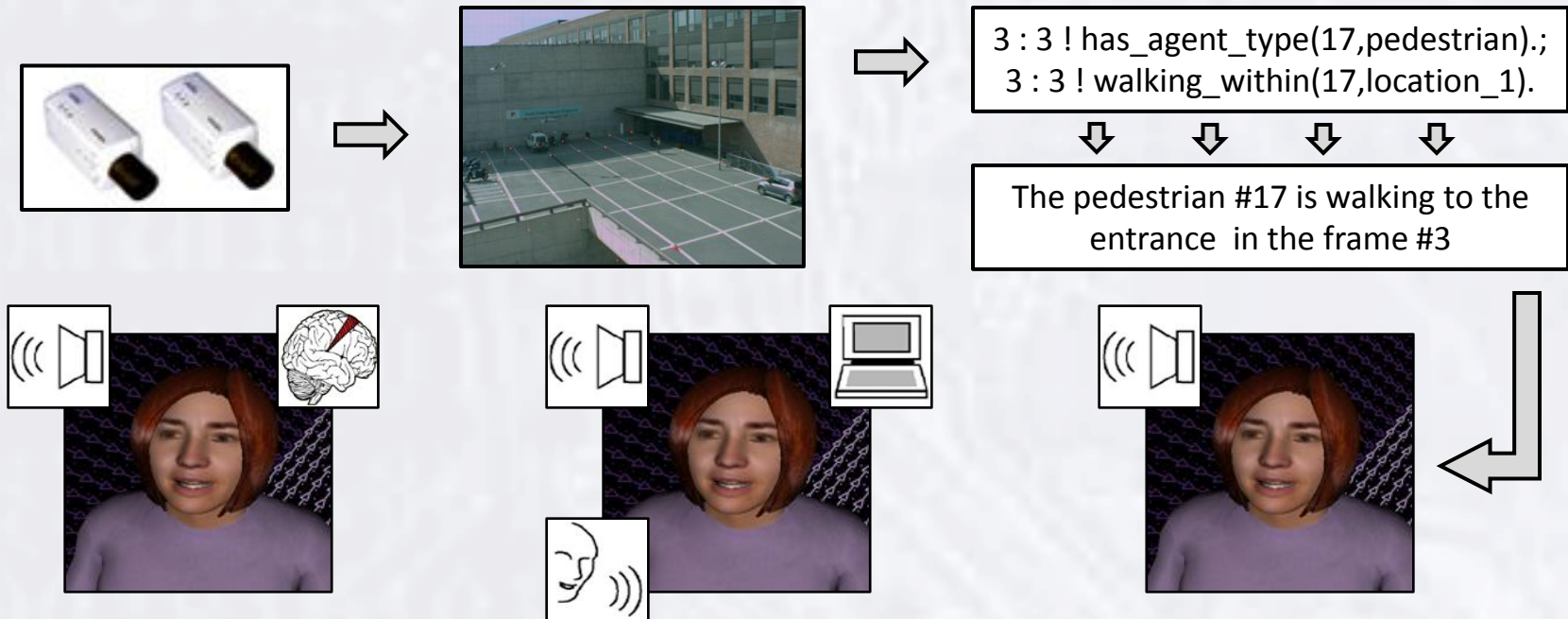
- **Hermes:** Sistema cognitivo computacional integrado por cámaras de vídeo y software que reconoce comportamiento humano y describe en lenguaje natural.



- **Mirando al mundo:** Análisis del comportamiento humano a partir de secuencias de vídeo capturadas con tres niveles de enfoque.

OBJETIVOS

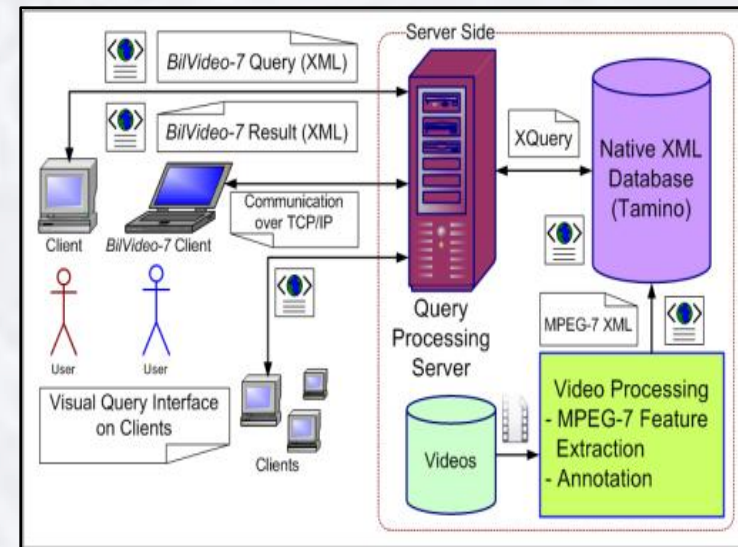
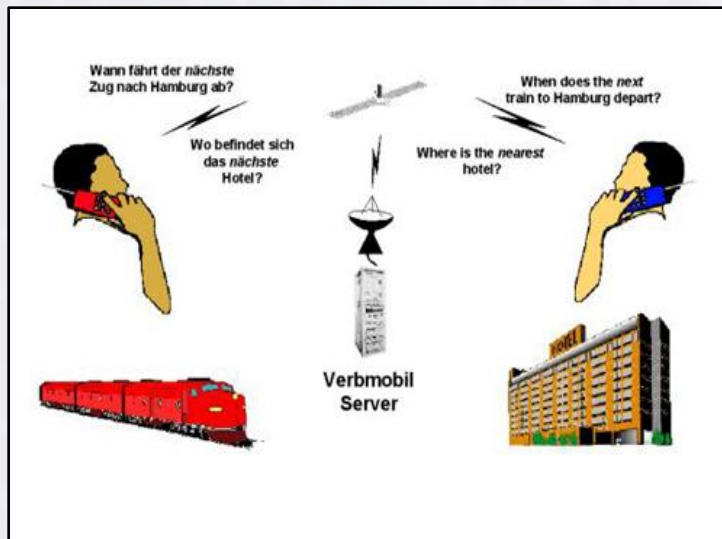
- ¿Qué ocurre?: El Avatar Virtual nos dirá lo que sucede en un área y atenderá a consultas sobre estos acontecimientos.



- **Finalidad:** Facilitar la tarea de búsqueda de sucesos ocurridos durante el día y prever aquellos acontecimientos con mayor probabilidad de repetirse .

ESTADO DEL ARTE

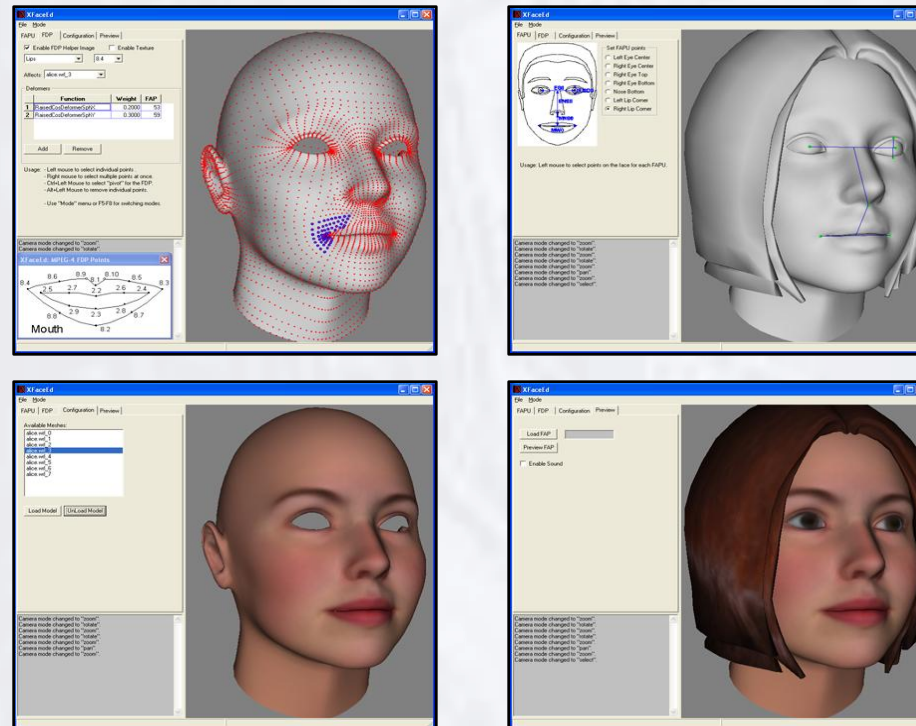
- Verbmobil: Sistema de traducción de voz a voz por teléfono.



- BilVideo: Sistema para realizar consultas sobre vídeos, mediante lenguaje natural.

BASE DEL PROYECTO

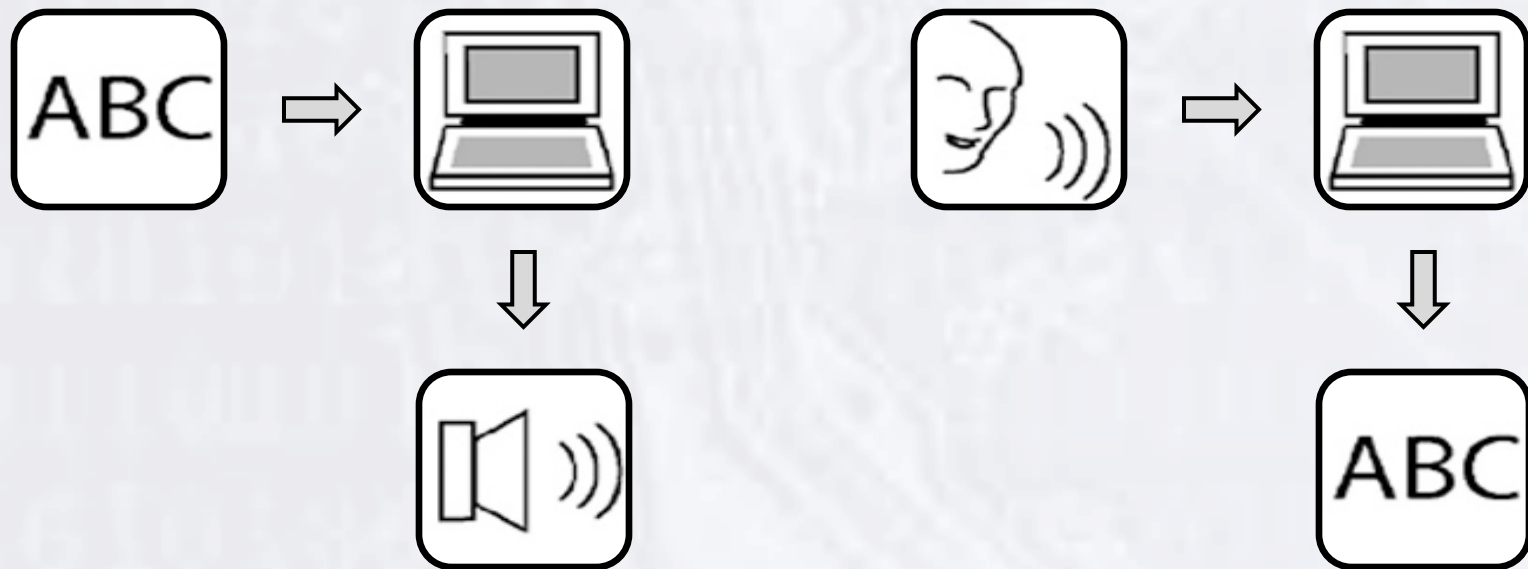
- XFace Toolkit: Herramienta básica para generar cabezas parlantes 3D.



- FA: Basado en animación facial de MPEG-4 que determina el grado de deformación.

BASE DEL PROYECTO

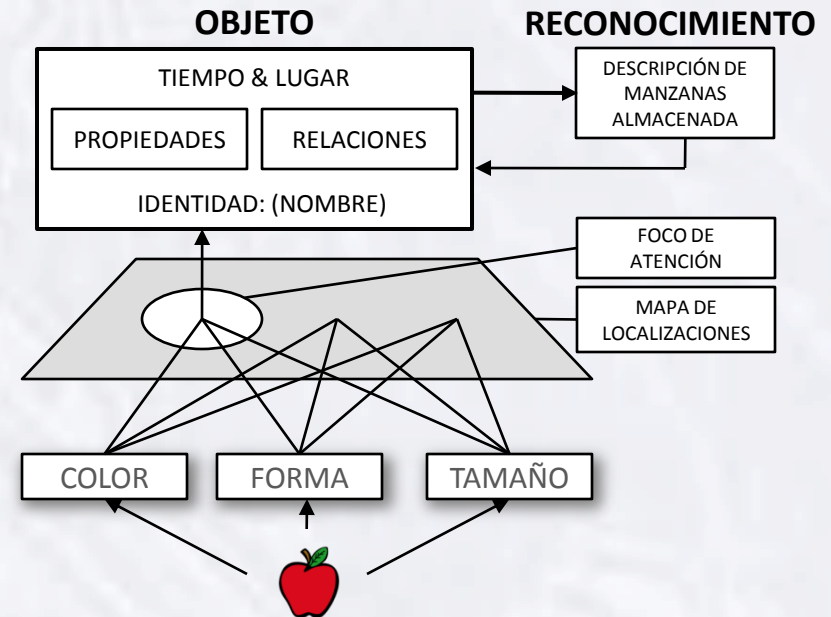
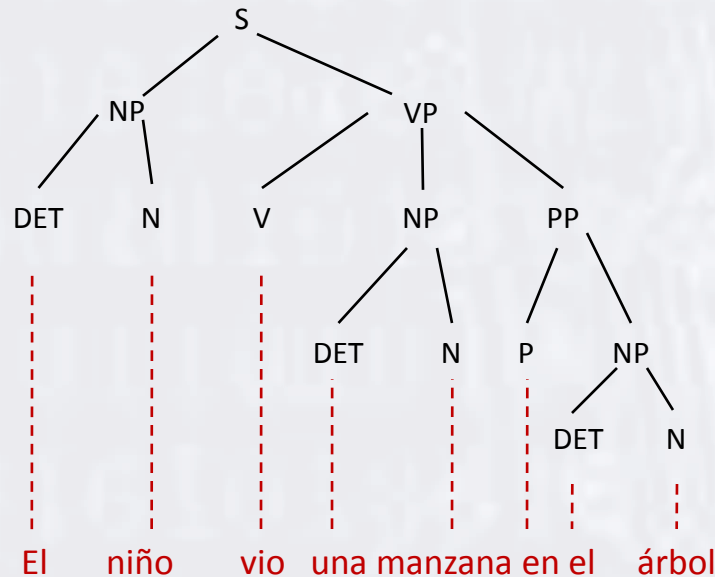
- Flite: Aplicación que convierte texto a voz.



- Julius: Aplicación que convierte voz a texto. Realizar consultas mediante la voz facilitan todavía más la comunicación hombre-máquina.

BASE DEL PROYECTO

- **PLN:** Procesamiento del lenguaje natural es una subrama de la inteligencia artificial y de la lingüística.



- **Finalidad:** Construir sistemas y mecanismos que permitan la comunicación entre personas y máquinas.

BASE DEL PROYECTO

- Tablas: Guardan información relevante sobre acontecimientos sucedidos.



- Inference: Almacena los predicados generados por el Sistema de Visión.



- Nltgresults: Contiene los mensajes en lenguaje natural.



- Nlu: Guarda las consultas que debe tratar el Avatar.



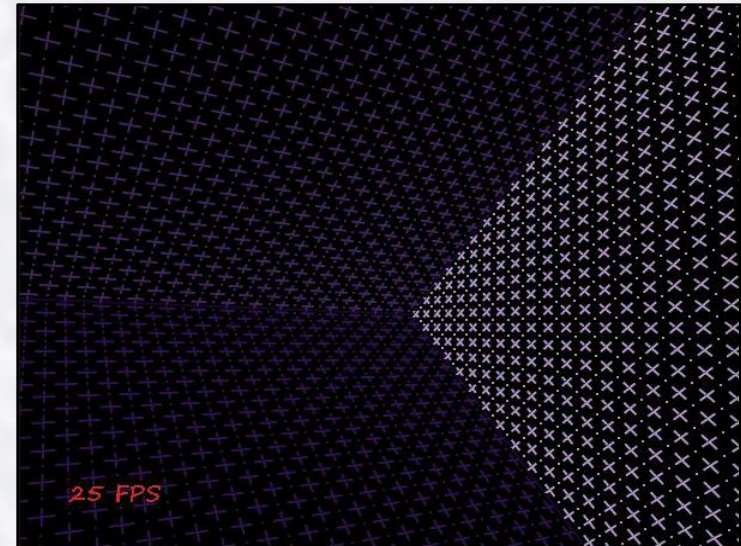
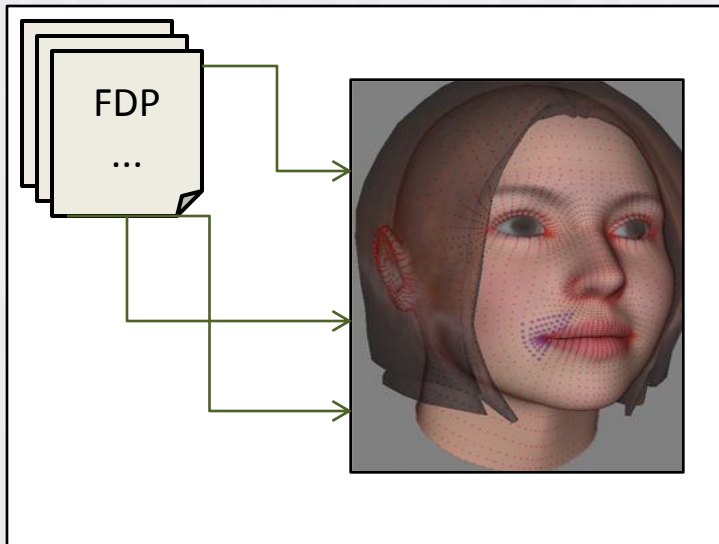
- Consciousness: Indica la prioridad de las acciones de los sucesos.



- Learn: Tiene en cuenta las acciones repetidas en un intervalo de tiempo.

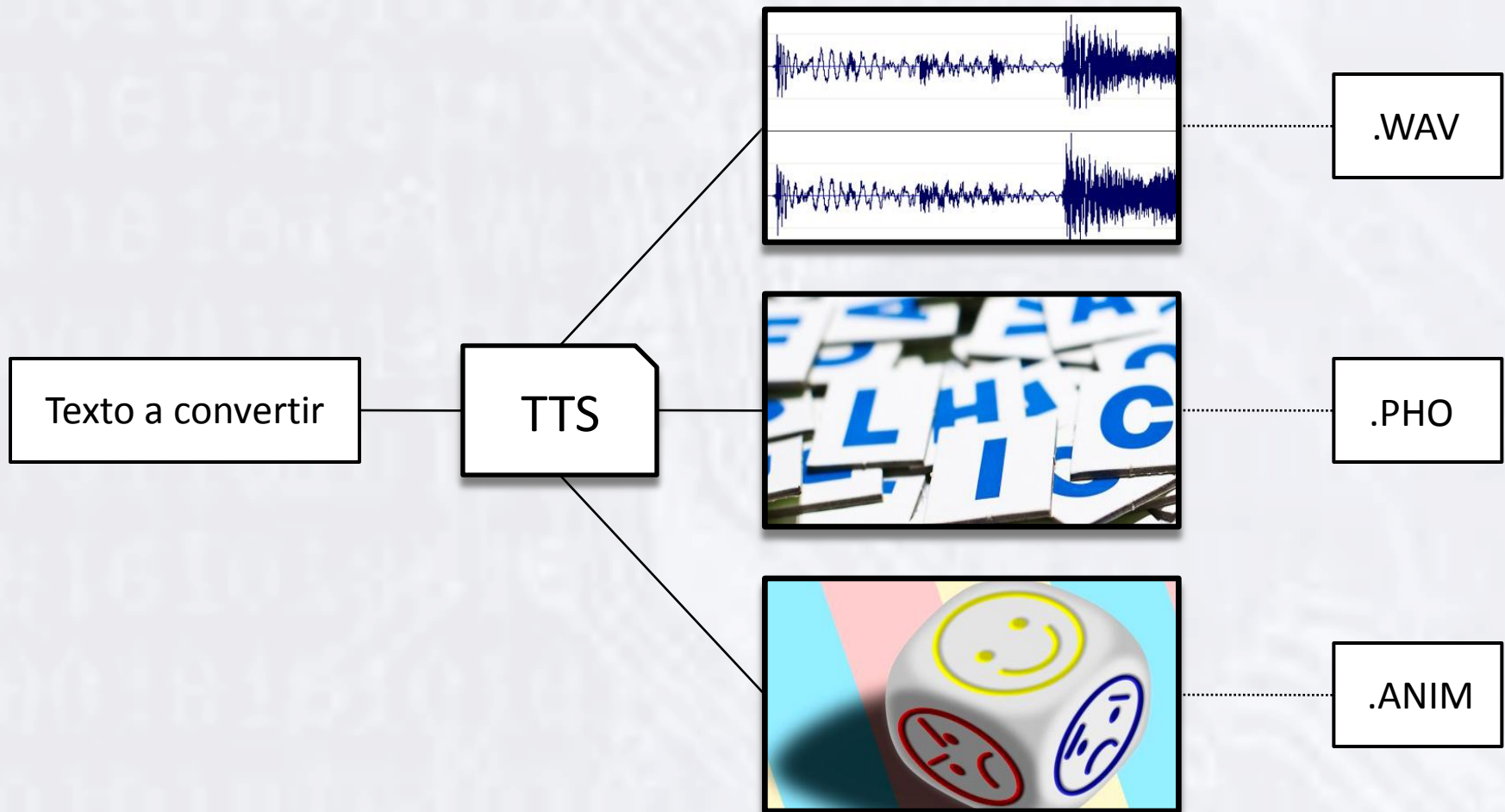
IMPLEMENTACIÓN

- FDP: Representación gráfica del fichero de salida de XFace.



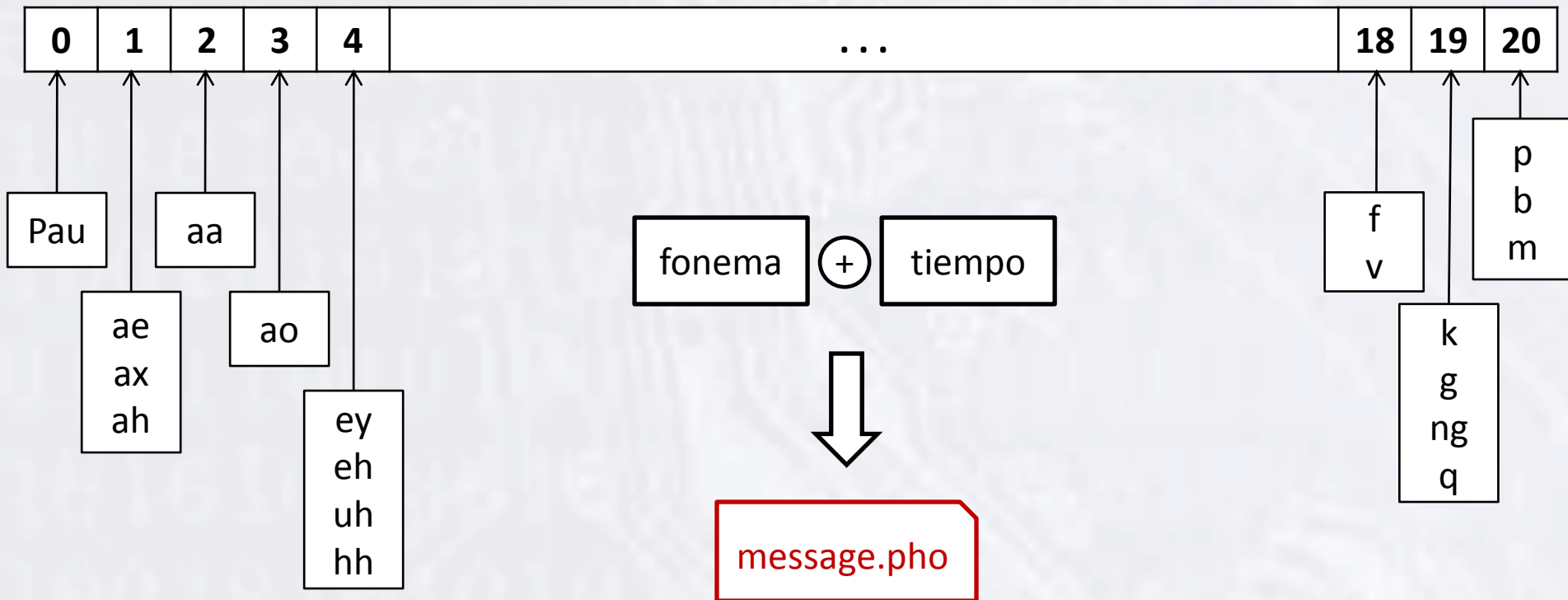
- Background: Cubo giratorio dinámico que envuelve al Avatar.

IMPLEMENTACIÓN



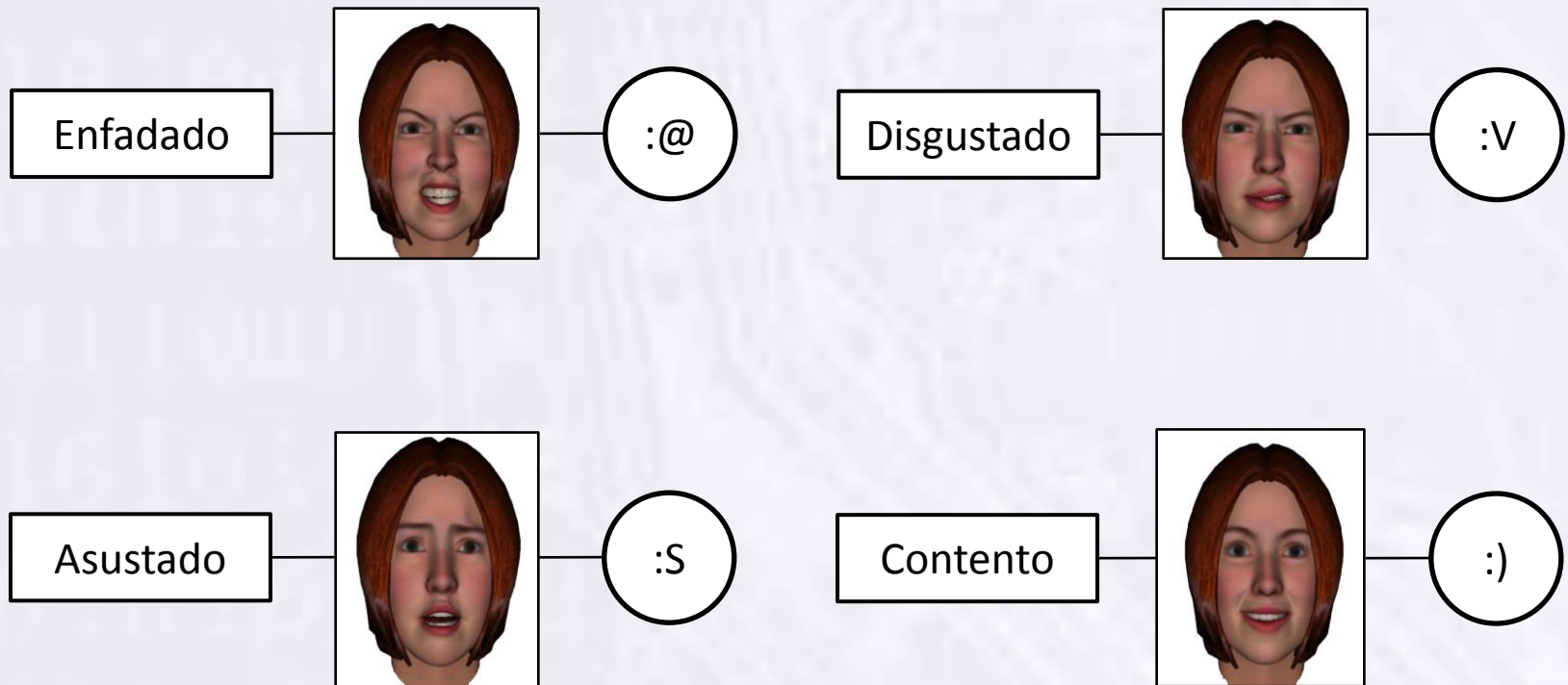
IMPLEMENTACIÓN

- Diccionario: Necesario para generar el fichero de fonemas.



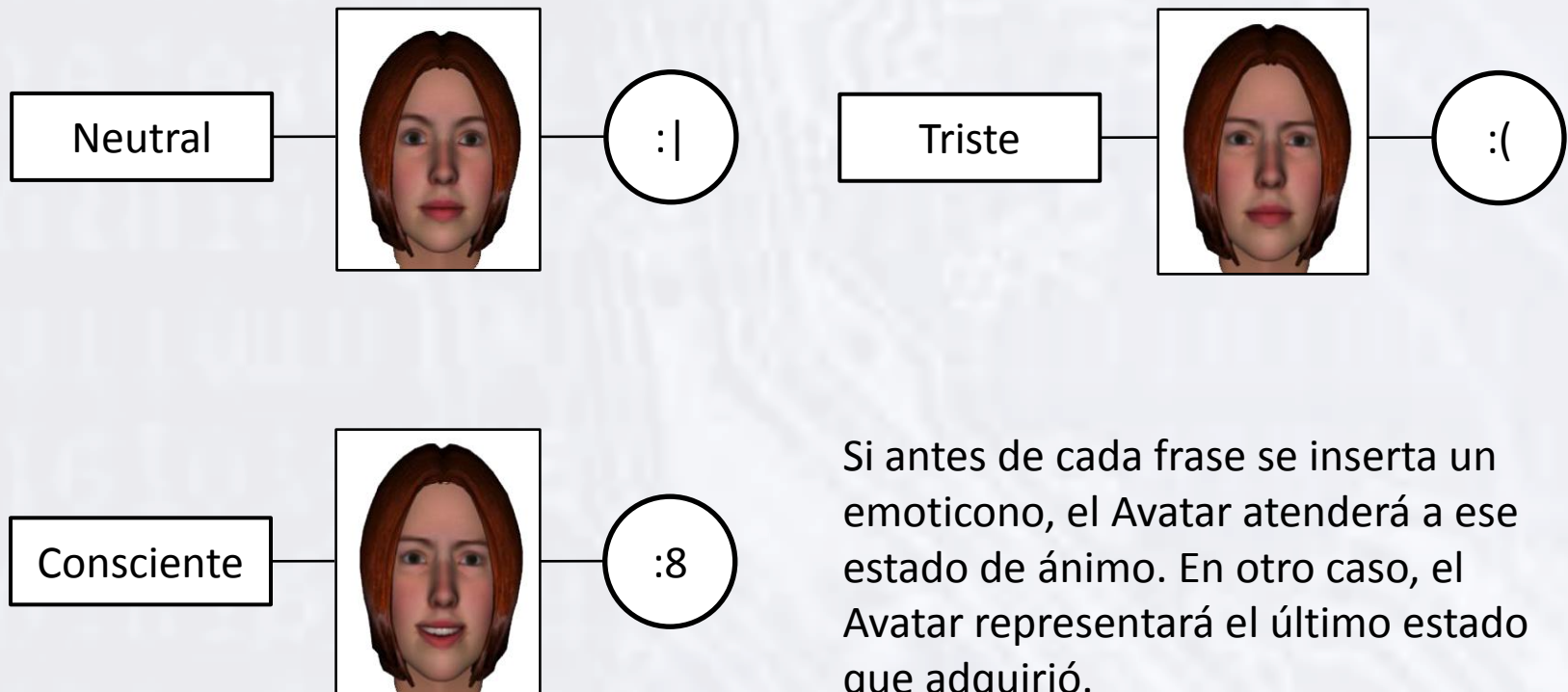
IMPLEMENTACIÓN

- Emoticonos: Ayudan a dar más realismo al Avatar cuando se expresa.



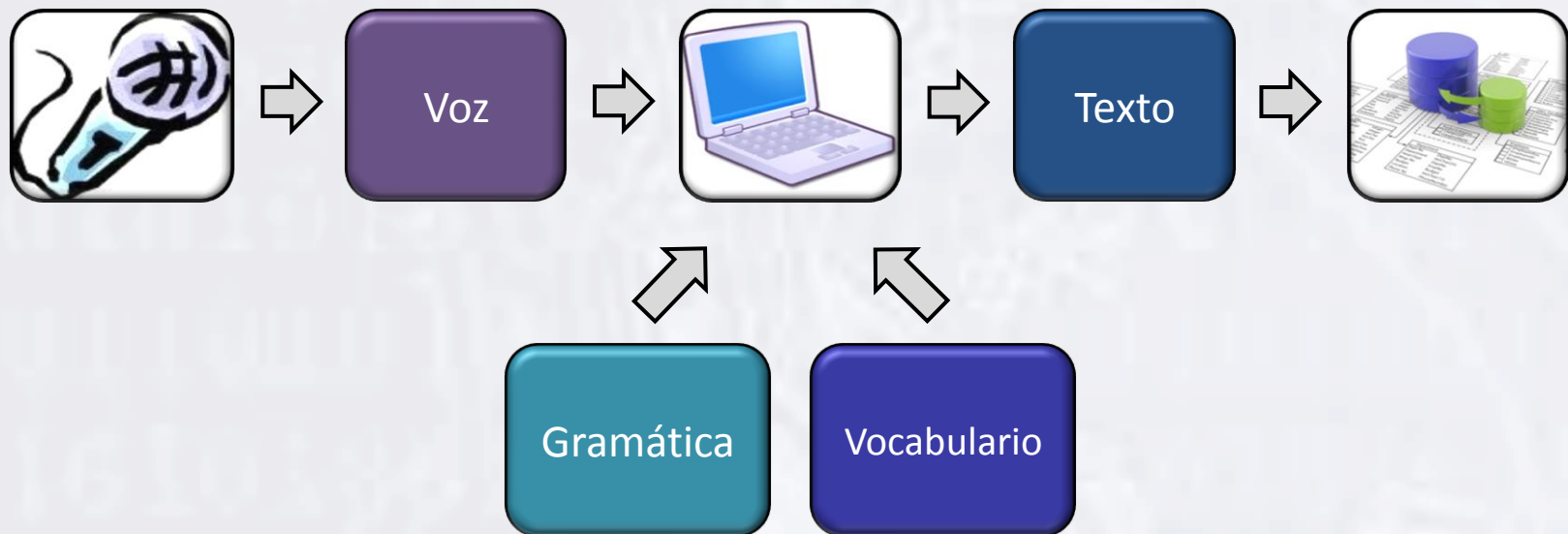
IMPLEMENTACIÓN

- Emoticonos: Ayudan a dar más realismo al Avatar cuando se expresa.



IMPLEMENTACIÓN

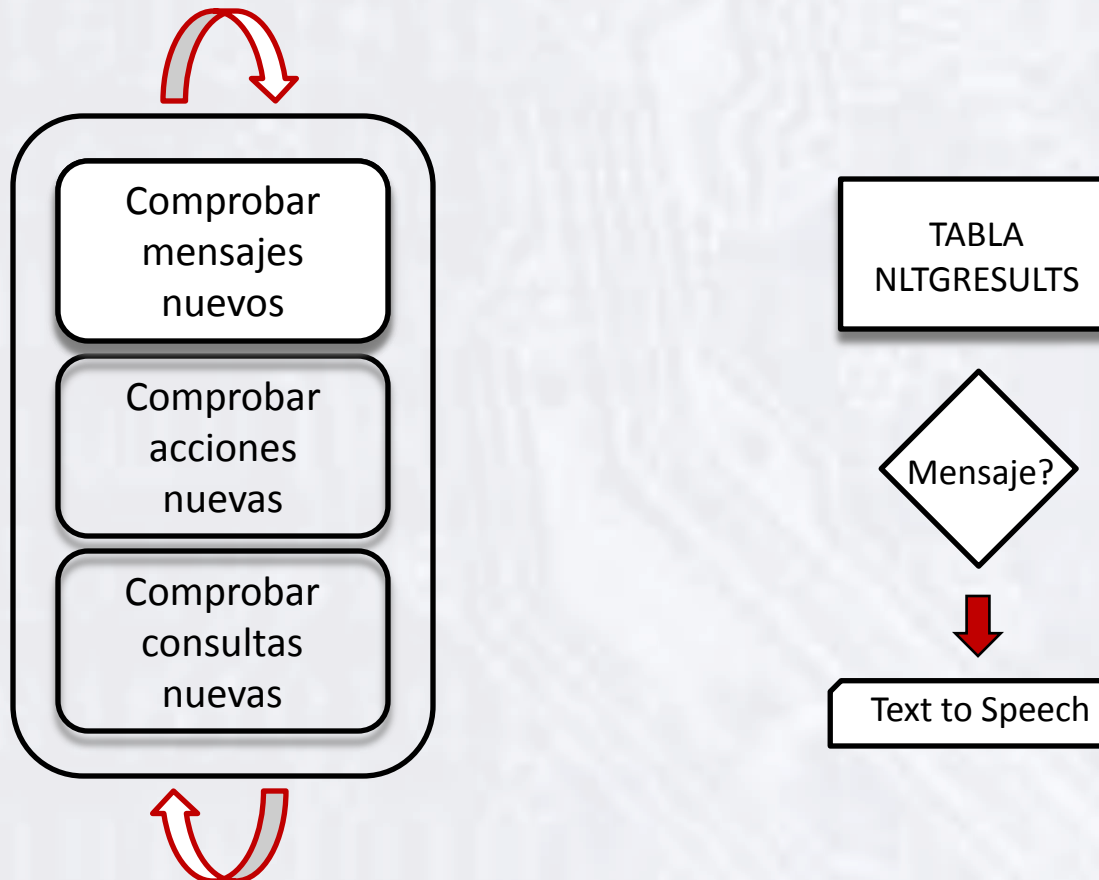
- Escuchando: La gramática determina qué frases reconocerá el Speech to Text.



- Petición: Una vez obtenida la consulta en modo texto, se le envía al analizador sintáctico para comprobar la estructura de ésta.

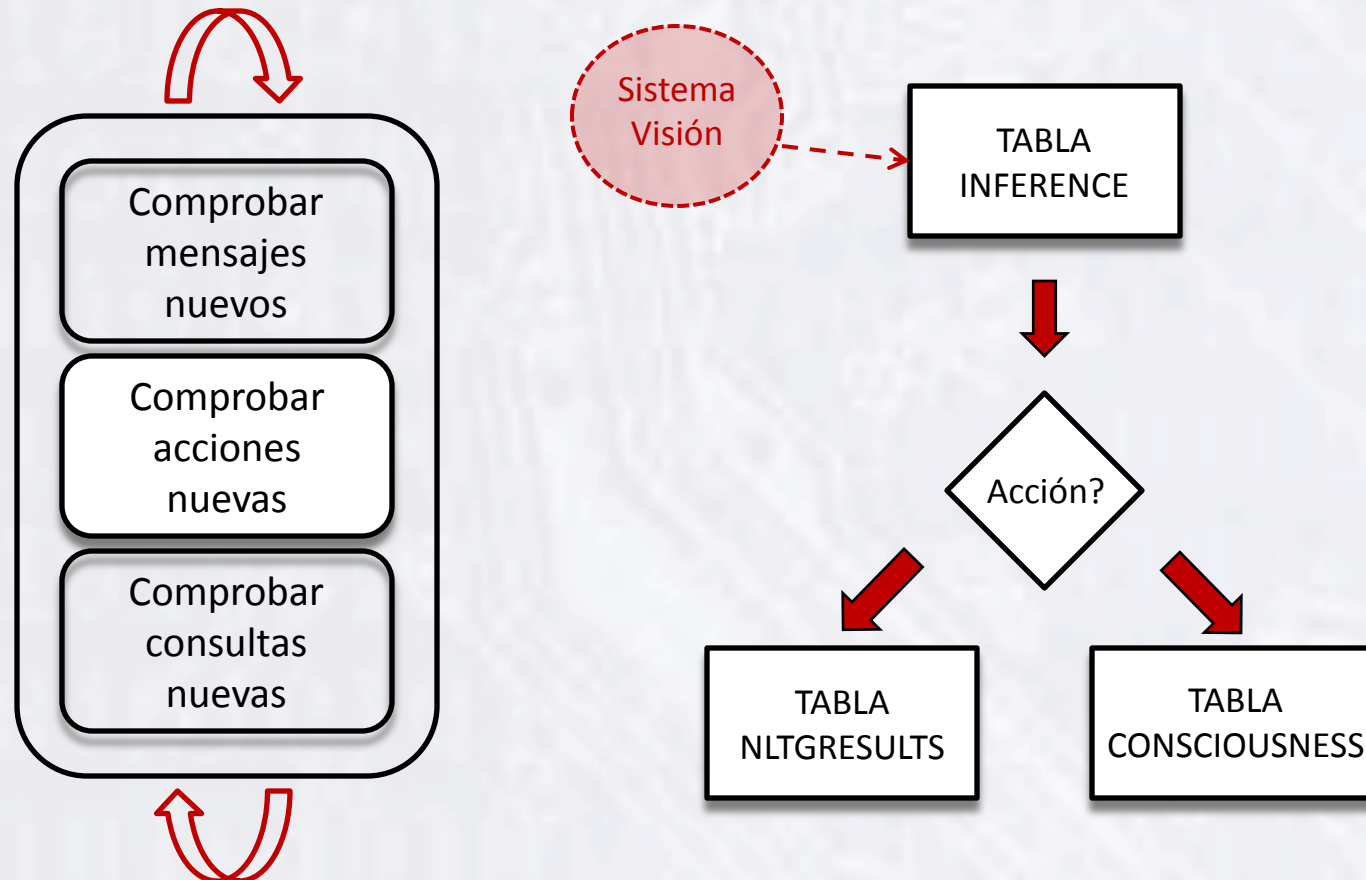
IMPLEMENTACIÓN

- Bucle infinito: Se comprueba constantemente si existe alguna entrada nueva.



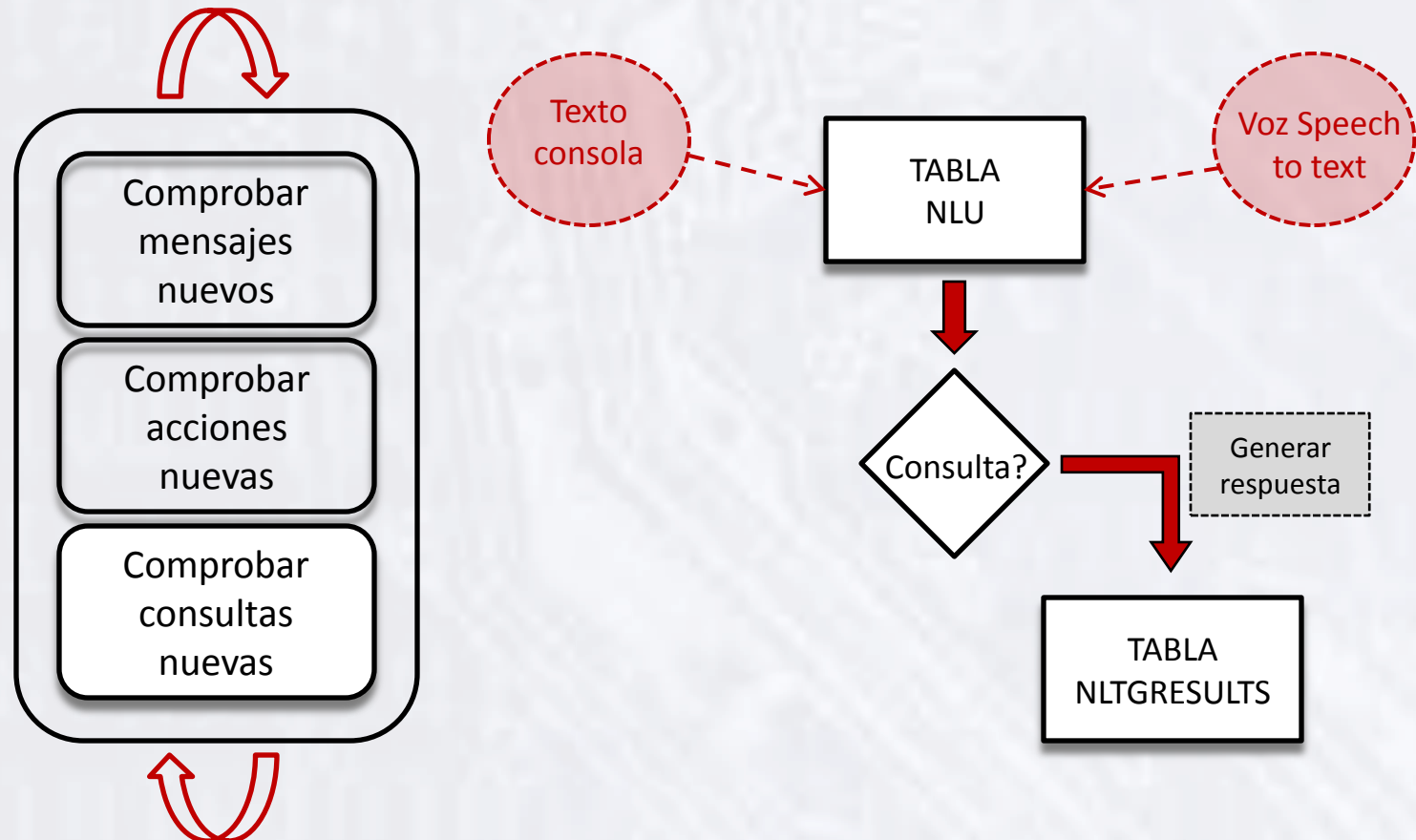
IMPLEMENTACIÓN

- Bucle infinito: Se comprueba constantemente si existe alguna entrada nueva.



IMPLEMENTACIÓN

- Bucle infinito: Se comprueba constantemente si existe alguna entrada nueva.



IMPLEMENTACIÓN

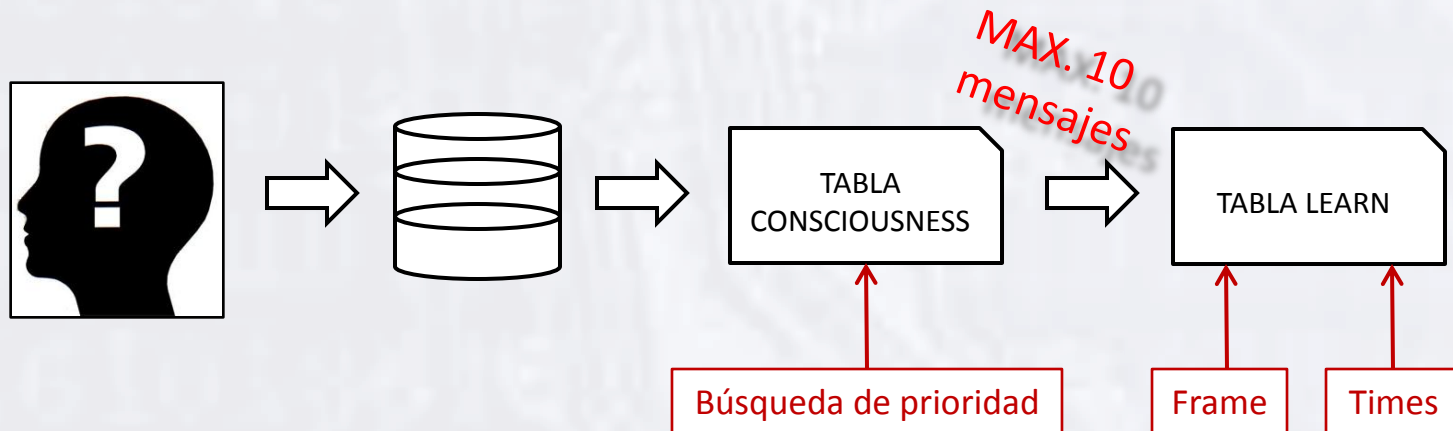
- **Consciencia:** Reconoce las acciones más importantes y las tiene en cuenta para determinar qué hechos son los más significativos del día.
 - Prioridad 1, para *theft, stole, stolen, steal, stealing*
 - Prioridad 2, para *hit, kick, punch, shove*
 - Prioridad 3, para *chase, chasing, run, running, escape*
 - Prioridad 4, para *meet, meeting, met, pick up, picking up*
 - Prioridad 5, para *leave, living, enter, entered*
 - Prioridad 6, para *walk, walking, walking with, walked*
 - Prioridad 7, para el resto de acciones



IMPLEMENTACIÓN

- **Consciencia:** Reconoce las acciones más importantes y las tiene en cuenta para determinar qué hechos son los más significativos del día.

- Query 7: “What is the most important thing happened today?”



- “There was a Action in the frame Frame”
- “The Agent # Action in the Location”

RESULTADOS

- Mensajes: Cuando el sistema recibe un nuevo suceso, lo convierte al lenguaje natural, lo envía a la tabla de mensajes y posteriormente lo reproduce con voz.



```

20 | 30 | :D the pedestrian 9 is meeting
19 | 30 | :D the pedestrian 7 is meeting
18 | 3 | :D the pedestrian 8 is meeting
17 | 30 | :D the pedestrian 3 is meeting
16 | 30 | :D the pedestrian 9 is meeting
15 | 30 | :D the pedestrian 7 is meeting
14 | 3 | :D the pedestrian 8 is meeting
13 | 42 | :S the pedestrian 8 is stole in the restricted area
12 | 42 | :| the child 67 is entered in the entrance
11 | 41 | :( the car 77 is chasing in the special area
10 | 41 | :( the car 11 is theft in the special area
9 | 40 | :( the pedestrian 12 is theft in the ramp
8 | 39 | :( the pedestrian 8 is theft in the highspeed area
7 | 33 | :D the pedestrian 8 is meeting in the highspeed area
6 | 30 | :D the pedestrian 3 is meeting in the near building wall
5 | 30 | :D the pedestrian 9 is meeting in the entrance
4 | 30 | :D the pedestrian 9 is meeting in the forbidden area
  
```



```

New Msg(5):List agents in the entrance
:| the pedestrian 0 is in the entrance
:| the pedestrian 9 is in the entrance
:| the pedestrian 1 is in the entrance
:| the pedestrian 8 is in the entrance
:| the pedestrian 10 is in the entrance
:| the pedestrian 2 is in the entrance
:| the pedestrian 3 is in the entrance
:| the pedestrian 6 is in the entrance
:| the pedestrian 0 is in the entrance
:| the pedestrian 9 is in the entrance
:| the pedestrian 1 is in the entrance
:| the pedestrian 8 is in the entrance
:| the pedestrian 1 is in the entrance
:| the pedestrian 2 is in the entrance
:| the pedestrian 3 is in the entrance
:| the pedestrian 6 is in the entrance
  
```

- Query: El Avatar atiende a las consultas generando un nuevo mensaje.

RESULTADOS

- Tabla Consciencia: Guarda las acciones con sus respectivas prioridades y el momento en el que se produjo el suceso.

23	11	chasing	41	3
22	10	theft	41	1
21	9	theft	40	1
20	8	theft	39	1
19	7	meeting	33	4
18	6	meeting	30	4
24	12	entered	42	5
25	13	stole	42	1
26	30	stolen	22	1
27	31	running	26	3
28	32	running	28	3
29	33	running	29	3
30	34	walking	29	6
31	35	walking	31	6
32	36	leaving	31	7
34	38	chasing	36	3

consciencia

```
New Msg(6):What is the most important thing happened today?
:8 There was a stolen in the frame 22
:S the pedestrian 3 is stolen in the forbidden area
:8 There was a stole in the frame 42
:S the pedestrian 8 is stole in the restricted area
:8 There was a stealing in the frame 36
:S the pedestrian 8 is stealing in the lowspeed area
:8 There was a running in the frame 29
:V the pedestrian 3 is running in the entrance
:8 There was a running in the frame 28
:V the pedestrian 8 is running in the special area
:8 There was a running in the frame 26
:V the pedestrian 3 is running in the lowspeed area
:8 There was a chasing in the frame 36
:( the pedestrian 90 is chasing in the forbidden area
```

consulta 7

- Query 7: “¿Qué es lo más importante sucedido hoy?”.

CONCLUSIÓN

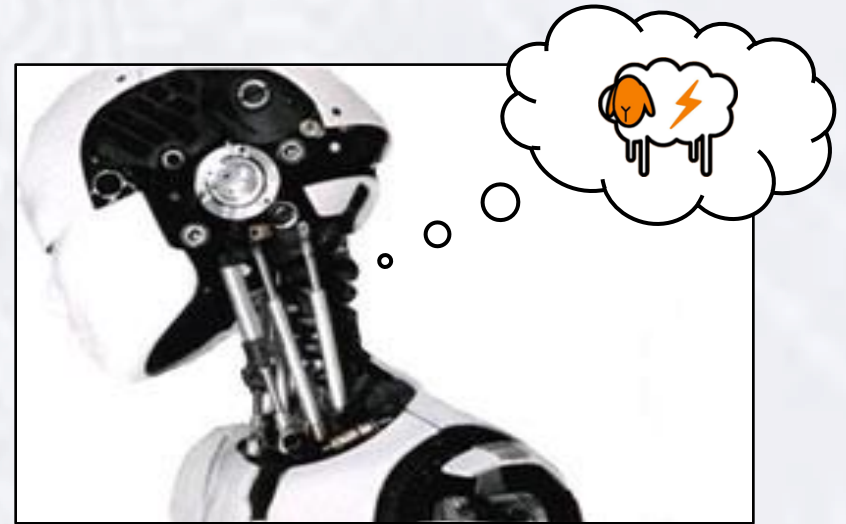
- **Eficacia:** No es necesario observar cientos de cámaras al mismo tiempo, ni todo el tiempo. Simplemente, hay que preguntarle al Avatar.



- **Atención:** Sistema de vigilancia y control de acontecimientos, sea cual sea el ámbito.

CONCLUSIÓN

- Vías de investigación: Trabajar más a fondo sobre la interacción con el Avatar. Qué éste amplíe su base del conocimiento automáticamente.



“Aprender a aprender”



GRACIAS POR SU ATENCIÓN