

Procesamiento Digital de Audio

Dr. Caleb Rascón
caleb@unam.mx

Procesamiento Digital de Audio

- El campo de señales de audio es normalmente considerado como el primito “especial” del área de Procesamiento de Señales.
 - Señales visuales son más “sexy”
 - Señales Biomédicas salvan vidas
- Mientras que procesar y analizar audio se considera una reliquia de los 90s.
 - MP3 y compresión de audio

Procesamiento Digital de Audio

- Dentro de una señal de audio hay un **mundo** enorme que ha sido poco explorado.
- No fue hasta recientemente (> 2006) que se ha analizado adecuadamente.
- Este **mundo** es lo que nos vamos a referir como la “Escena Auditiva”.
 - Y es muy compleja.

¿No me creen?

Experimento con Audio

Cierren los ojos

Y

Sólo pónganle atención a la primera voz que
escuchen

Complejidad de Escena Auditiva

- ¿Qué pudieron hacer con sólo audio?
 - Saber dónde están las varias fuentes de sonido.
 - Separar las fuentes en canales con sólo una fuente.
 - Saber qué es cada una de esas fuentes y etiquetarlas.
 - Filtrar el sonido de las fuentes etiquetadas “ruido”.
 - Saber qué es lo que está diciendo cada una de las fuentes del tipo “persona”.
 - Decidir cuál es la más importante “ponerle atención”.
 - Comprender qué es lo que está diciendo la fuente más importante.
 - Reaccionar a ruidos no esperados.
 - ... en tiempo real, en ambientes muy dinámicos.

¿Para qué?

Motivación

- Audio está en todo nuestro alrededor.
- El límite de rango del que se puede recibir una señal de audio es alto:
 - 360 grados ambos vertical y horizontalmente.
- Comparado a:
 - Visión: es ~ 170 grados horizontalmente y ~ 90 grados verticalmente.
 - Olfato: rango comparable, pero sin dirección.
 - Tacto: requiere movimiento del sensor.

Motivación

- Hay mucha información que puede ser extraída de una señal de audio:
 - Dirección y distancia de una fuente de sonido
 - Reconocimiento de clase de fuente de sonido
 - Reconocimiento del habla
 - Tamaño del entorno
 - Materiales de paredes

Motivación → Desafío

- Toda esa información de un arreglo de datos de una dimensión.
- Si tenemos varios micrófonos, podemos tener varios arreglos que analizar.
- ... en línea/tiempo real.

Motivación → Desarrollo

- Práctica de primera mano en desarrollar software de procesamiento.
- Hoy en día, si estás procesando señales, lo estás haciendo con una computadora.
- Lenguajes variantes de C son los más prácticos en utilizar, por su robustez, eficiencia de manejo de recursos, así como su precio.

Motivación → Telecomunicaciones

- Si consideramos al micrófono como un tipo de antena, muchos de los conceptos de telecomunicaciones se pueden aplicar en procesamiento de señales de audio.
 - Filtrado
 - Comparación de señales
 - Beamforming

Motivación → Aplicaciones

Audición Robótica

Audición Robótica

- Varias definiciones.
- La que más me gusta:

“Rama que tiene como objetivo otorgar a un ente no-humano una capacidad auditiva cercana a la de un humano oyente.”

Audición Robótica

- Varias definiciones.
- La que más me gusta:

“Rama que tiene como objetivo otorgar a un **ente no-humano** una capacidad auditiva cercana a la de un humano oyente.”

- Usualmente un robot de servicio.

Motivación → Audición Robótica

Videos

Objetivo del Curso

Exponer al(la) alumno(a) a temas de procesamiento de señales de audio, donde se cubrirán desde los conceptos teóricos hasta los aspectos de implementación.

Intención del Curso

Al terminar este curso, le sea posible al(la) alumno(a) crear, de una manera eficiente, software que analice, procese, y regrese resultados de señales de audio en línea.

Suposición del Curso

Aunque se espera que el(la) alumno(a) tenga bases previas de Señales y Sistemas, así como de programación, el curso se llevará a cabo de tal manera que dichas bases serán refinadas y repasadas.

Intención del Curso

Al terminar este curso, le sea posible al(la) alumno(a) crear, de una manera eficiente, **software** que analice, procese, y regrese resultados de señales de audio en línea.

Desarrollo de Software

- Vamos a programar... mucho.
- En el lenguaje C.
 - Utilizando *struct*'s y dobles apuntadores.
 - Vamos a compilar directamente en línea de comando.
 - Sin IDE.
- También vamos a prototipar en Octave.
 - Alternativa de código libre de Matlab.
- Por lo tanto, necesitaremos computadoras.

¿Todos tenemos acceso a alguna computadora?

¿Todos tenemos acceso a alguna computadora?

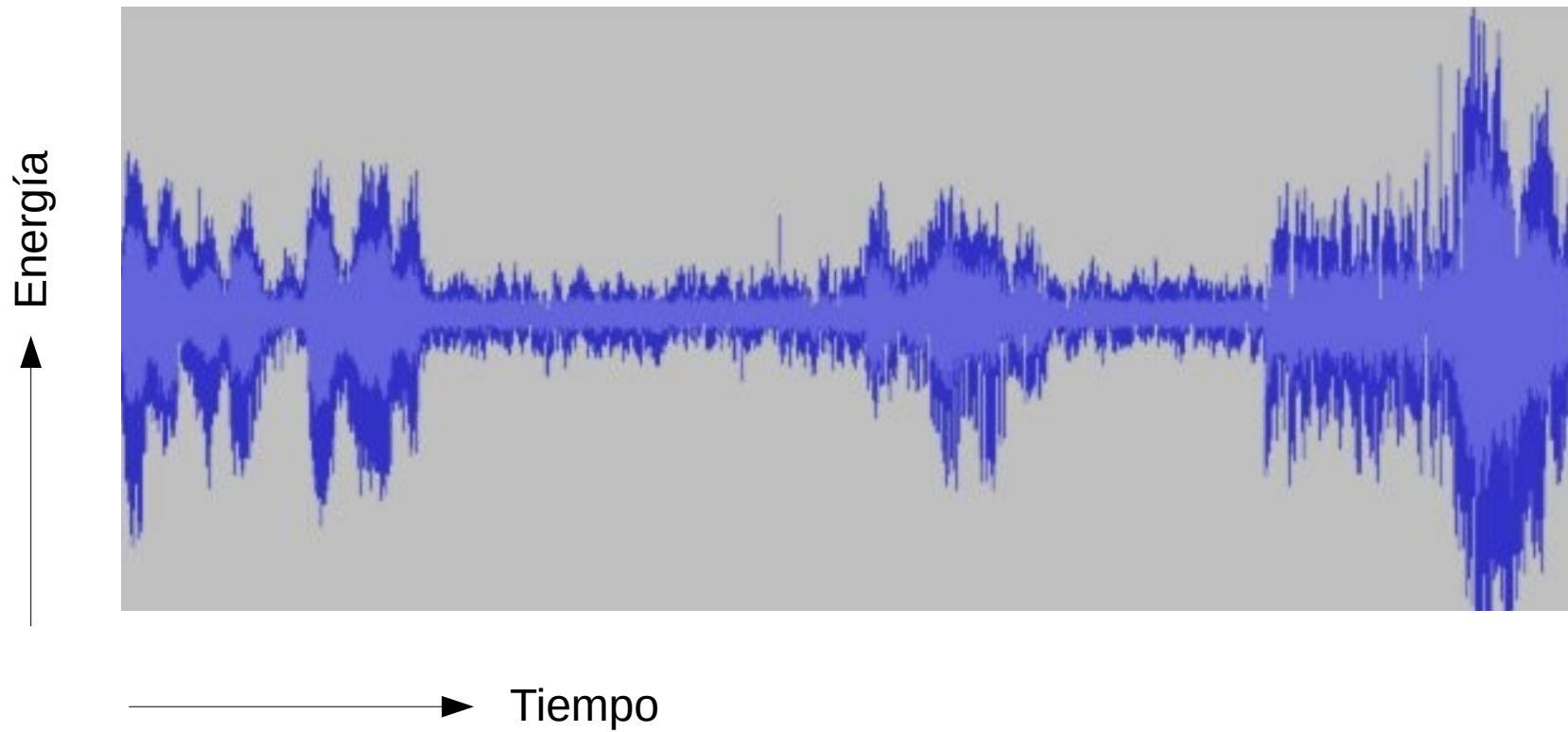
Con Linux instalado, nativamente...

Se puede virtualizar, pero se han tenido malas experiencias.

Intención del Curso

Al terminar este curso, le sea posible al(la) alumno(a) crear, de una manera eficiente, software que analice, procese, y regrese resultados de **señales de audio** en línea.

Señales de Audio



Señal de Audio

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0.1	0.0	0.2	0.4	0.7	1.1	1.0	0.8	0.6	0.6	0.3	0.2	0.1	0.0	0.0	0.2	0.5	0.4

Arreglo de valores, en el que cada celda es un momento en el tiempo, y el valor de la celda es el valor de energía.

Intención del Curso

Al terminar este curso, le sea posible al(la) alumno(a) crear, de una manera eficiente, software que analice, procese, y regrese resultados de señales de audio **en línea**.

“En línea” != “En tiempo real”

“En línea” \sim “En tiempo real”

En Línea

- Lo contrario a “fuera de línea”.
- El procesamiento se lleva a cabo durante la captura, no después.

En Tiempo Real

- Implica “en línea”.
- Realmente no hay una definición definitiva.
 - Difiere dependiendo del área de aplicación.
- Para propósitos de este curso:
 - “Tan rápido como un humano habla por un micrófono y se escucha a si mismo por una bocina”.

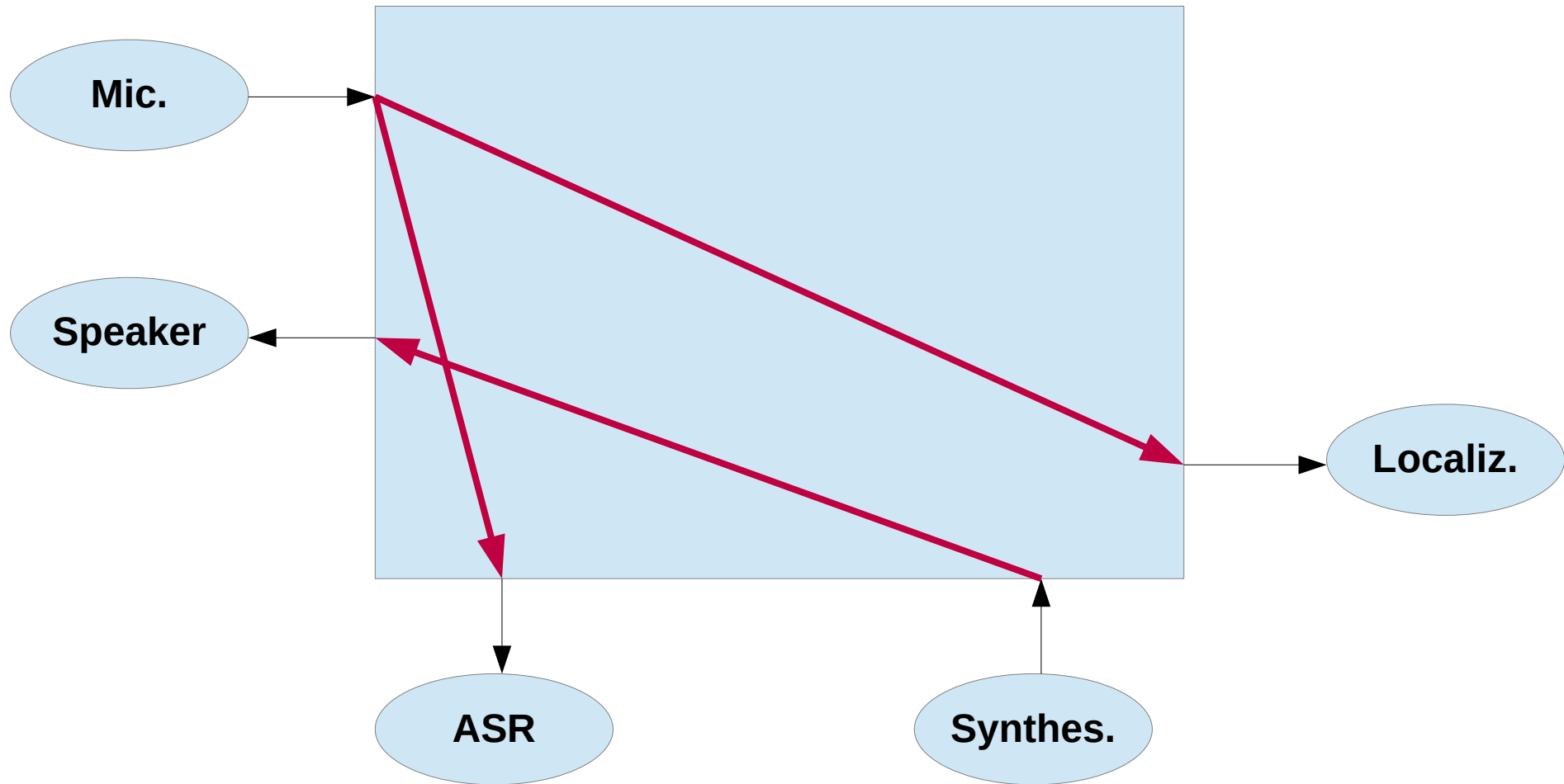
Independientemente...

- Nos haremos muy buenos amigos de una biblioteca de procesamiento de señales de audio llamada:
 - **JACK Audio Connection Kit**
- Su documentación presume hacer procesamiento de audio en tiempo real, pero lo que nos interesa es que lo hace en línea.
- Pero, tiene otras facilidades:

JACK

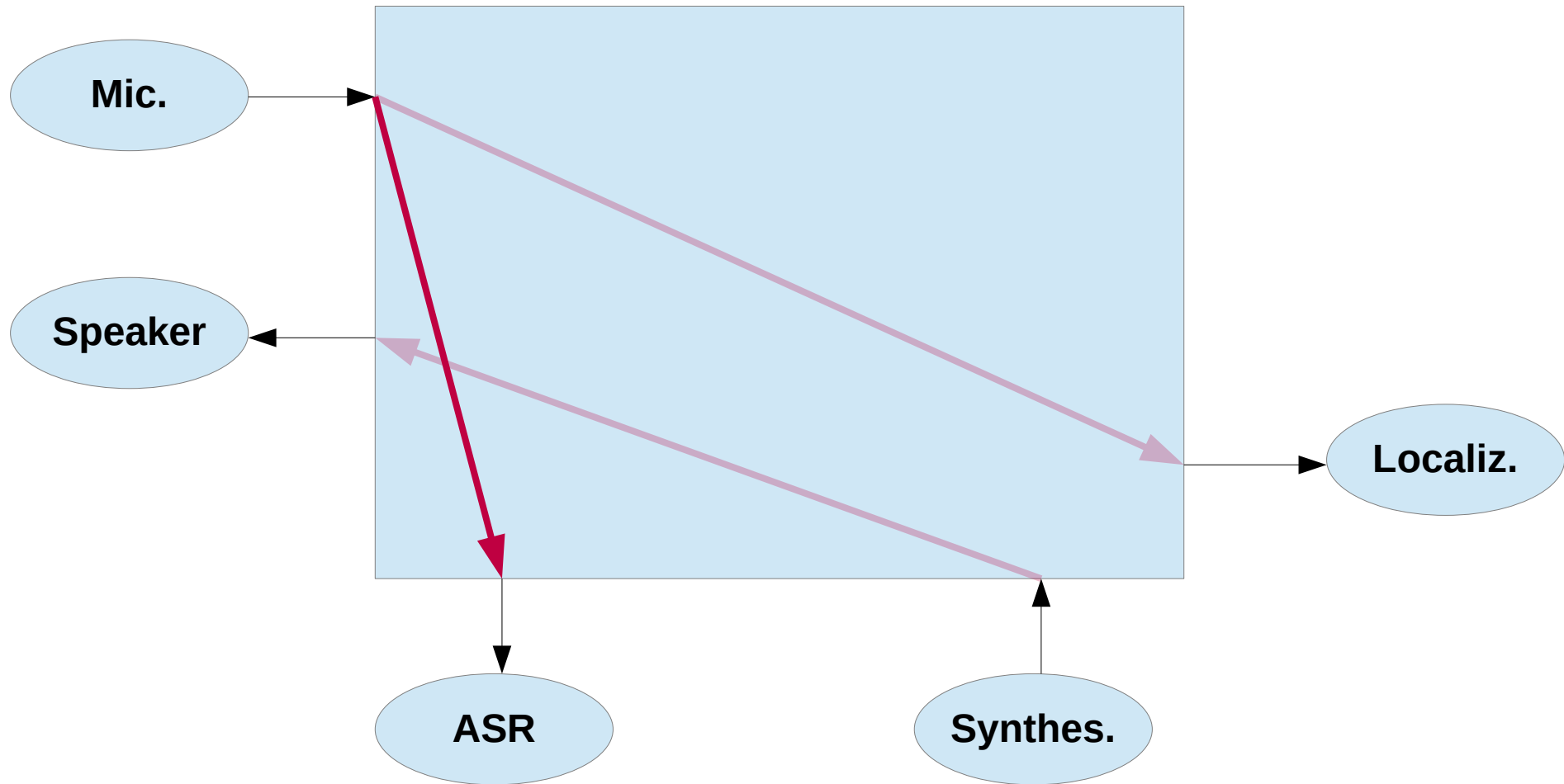
- Otorga la facilidad de crear agentes de JACK que se pueden conectar al servidor de JACK, el cual, a su vez, está conectado a dispositivos de audio (bocinas para salida, micrófonos para entrada, etc.).

JACK



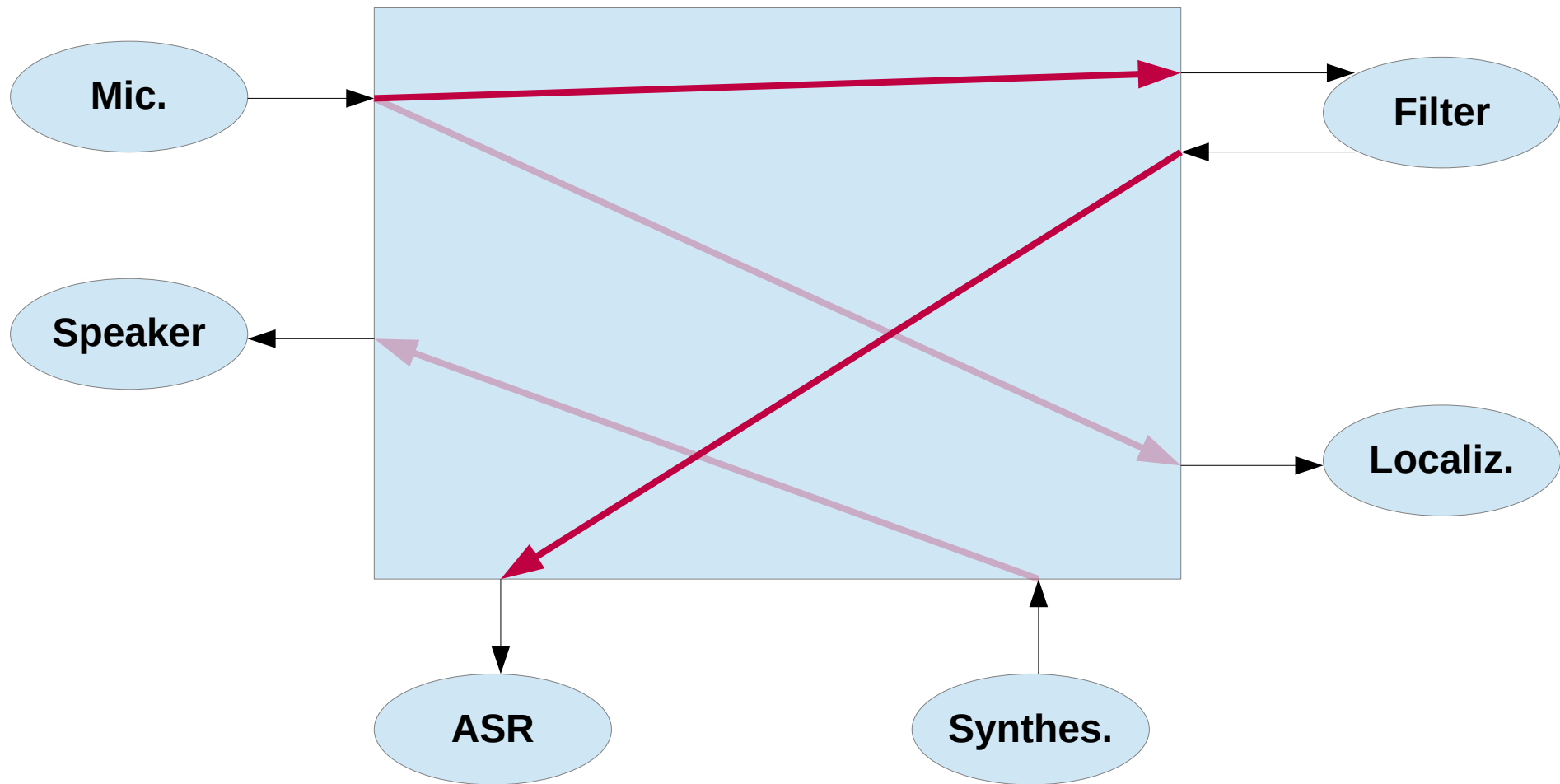
JACK

ASR en necesidad de filtro



JACK

ASR en necesidad de filtro

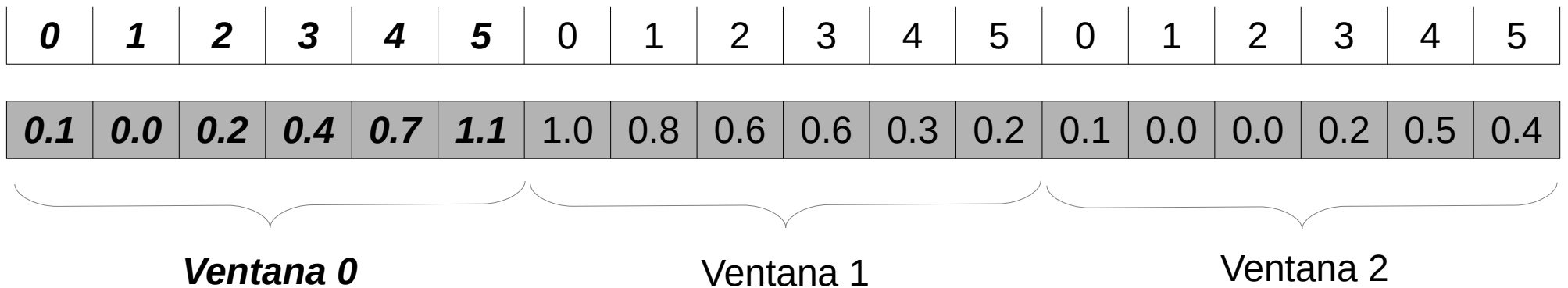


JACK

- Cada agente se comunica con JACK por medio de escribir y/o leer valores de energía en arreglos de datos.
- Dichos arreglos representan *ventanas* de audio.
 - También conocidos como “periodos”.

En Línea → JACK

Ventanas de tamaño: 6



JACK

- De esta manera, mientras escribamos/leamos adecuadamente dichos arreglos de datos, JACK se encarga de entregar los datos a tiempo ya sea a:
 - Los dispositivos
 - Otros agentes

Procesamiento Digital de Audio

Al final de este curso, deben de poder llevar a cabo un análisis de **alguna parte de la escena auditiva.**

Partes de la Escena Auditiva

- Saber dónde están las diversas fuentes de sonido alrededor.
- Separar las fuentes en canales con sólo una fuente.
- Saber qué es cada una de esas fuentes y etiquetarlas.
- Filtrar el sonido de las fuentes etiquetadas “ruido”.
- Saber qué es lo que está diciendo cada una de las fuentes del tipo “persona”.
- Decidir cuál es la más importante “ponerle atención”.
- Comprender qué es lo que está diciendo la fuente más importante.
- Reaccionar a ruidos no esperados.

Partes de la Escena Auditiva

- **Saber dónde están las diversas fuentes de sonido alrededor.**
- **Separar las fuentes en canales con sólo una fuente.**
- Saber qué es cada una de esas fuentes y etiquetarlas.
- Filtrar el sonido de las fuentes etiquetadas “ruido”.
- Saber qué es lo que está diciendo cada una de las fuentes del tipo “persona”.
- Decidir cuál es la más importante “ponerle atención”.
- Comprender qué es lo que está diciendo la fuente más importante.
- Reaccionar a ruidos no esperados.

Realmente...

- Este curso les dará las bases para poder hacer proyectos de procesamiento de audio muchísimo más complejos.
 - Extrayendo otras partes de la escena auditiva.

cough tesis *cough*

Página del Curso

<http://calebrascon.info/PDA>

Curso

- Se mantendrá un balance entre la teoría y la práctica.
 - Primer mes, repaso de las bases de:
 - *Lenguaje C / Bash
 - Captura de Audio
 - Señales y Sistemas (*Transformada de Fourier*)
 - Resto:
 - Filtrado y Detección de Actividad de Voz
 - Revisión de algoritmos para:
 - Estimación de Dirección de Arribo
 - Separación de Fuentes en Línea
 - Implementación de dichos algoritmos

Curso

- Tareas prácticas.
- Un examen parcial al fin del primer mes.
- Un proyecto final.
 - Con un periodo extenso al final del semestre para desarrollarlo.
 - Se hará una presentación de avance a la mitad de este periodo para ver cómo van.
 - Se hará una presentación final al término de este periodo con el cual se evaluará el semestre.

Proyecto Final

- DESEABLE. *Resolver ambos:*
 - Estime la dirección de las fuentes en el ambiente.
 - Las separe en diferentes canales.

- Máxima calificación: 10
... no es para nada trivial.

Proyecto Final

- MÍNIMO. *Resolver uno:*
 - Estime la dirección de las fuentes en el ambiente.
 - *O, asumiendo que ya se conocen las direcciones,* las separe en diferentes canales.
- Máxima calificación: 9

Proyecto Final

- Presentación de 20 minutos, describiendo:
 - Algoritmo(s) utilizado(s) y por qué.
 - Problemas que se encontraron.
 - Si se pudieron resolver: ¿cómo?
 - Si no se pudieron resolver: potenciales soluciones.
 - Demostración del sistema, en vivo.

Información de Contacto

- Mi oficina: IIMAS, 4o piso, Oficina 403.
- Mi correo:
caleb@unam.mx
caleb.rascon@gmail.com