

Métricas de mimetización acústico-prosódica en hablantes y su relación con rasgos sociales de diálogos

Juan Manuel Pérez

21 de marzo de 2016

¿Cómo dijo?

- Sistemas de diálogo humano-computadora son cada vez más frecuentes, y sus aplicaciones comprenden una amplia gama de rubros

¿Cómo dijo?

- Sistemas de diálogo humano-computadora son cada vez más frecuentes, y sus aplicaciones comprenden una amplia gama de rubros
- Bien en la dimensión lingüística, mal en todo lo superestructural: emociones, actitudes, intenciones.

¿Cómo dijo?

- Sistemas de diálogo humano-computadora son cada vez más frecuentes, y sus aplicaciones comprenden una amplia gama de rubros
- Bien en la dimensión lingüística, mal en todo lo superestructural: emociones, actitudes, intenciones.
- Mimetización: Fenómeno inconsciente que se manifiesta a través de la adaptación de los hablantes. Fuertemente emparentada con el sentimiento de empatía.

¿Cómo dijo?

- Sistemas de diálogo humano-computadora son cada vez más frecuentes, y sus aplicaciones comprenden una amplia gama de rubros
- Bien en la dimensión lingüística, mal en todo lo superestructural: emociones, actitudes, intenciones.
- Mimetización: Fenómeno inconsciente que se manifiesta a través de la adaptación de los hablantes. Fuertemente emparentada con el sentimiento de empatía.
- Objetivo del trabajo: Explorar y refinar una métrica de la mimetización acústico-prosódica, y validar que capture ciertas percepciones sociales en un corpus de diálogos en inglés.

1 Sistemas actuales

- 1 Sistemas actuales
- 2 Bien en la parte lingüística de la comunicación: entender y transmitir mensajes estructuralmente correctos.

- 1 Sistemas actuales
- 2 Bien en la parte lingüística de la comunicación: entender y transmitir mensajes estructuralmente correctos.
- 3 Mal en la parte superestructural: intercambio de emociones, actitudes, etc.

- 1 Sistemas actuales
- 2 Bien en la parte lingüística de la comunicación: entender y transmitir mensajes estructuralmente correctos.
- 3 Mal en la parte superestructural: intercambio de emociones, actitudes, etc.
- 4 El presente trabajo trata de hacer un (pequeño) aporte sobre el análisis de la “naturalidad” de las conversaciones.

Ejemplos de “falta de naturalidad”

- ① Sistemas de llamadas comerciales
- ② Siri, Google Now
- ③ Otros?



- El “cómo” decimos las cosas, a diferencia del “qué”
- Parte fundamental del mensaje oral
- Algunas características que la definen: acentuación, velocidad, tono, ritmo, volumen.
- Es justamente en lo principal que fallan los sistemas humano-computadoras hoy día

- ① Fenómeno también conocido como mimetización, convergencia, efecto camaleón, etc.
- ② Consiste en la adaptación que ocurre entre hablantes a varios niveles: sintáctico, prosódico, en las posturas, etc.
- ③ Fenómeno ubícuo e inconsciente en la comunicación humana

Recuerden verificarlo en la próxima conversación que tengan

¿Y cómo lo medimos?

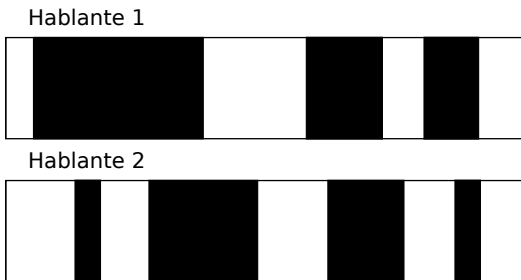
- La definición de entrainment hasta acá vista es muy subjetiva ¿Cómo definimos una medida para esto?
- Vamos a explorar una métrica definida en trabajos anteriores, pulirla un poco, y verificar que efectivamente capture ciertas características del entrainment.
- ¿Cómo? Aplicándola a un corpus con anotaciones sociales, y verificando la relación entre las percepciones sociales y la métrica del *entrainment*

Otras métricas

de entrainment prosódico

- La mimetización es un fenómeno tanto lineal: se va acentuando a lo largo de la conversación
- Pero también es un fenómeno dinámico: va variando localmente a lo largo de la conversación.
- Muchas métricas sólo toman la parte global, dividiendo la conversación en 2 o más partes y luego calculando la diferencia entre las medias de las diferentes variables acústicas en cada sección.
- Otro problema que tienen algunas métricas es que no son automáticas: requieren de anotaciones manuales sobre las conversaciones, por ejemplo patrones de entonación.

Problema del alineamiento de tiempo



Uno de los problemas que tenemos a la hora de construir métricas de entrainment

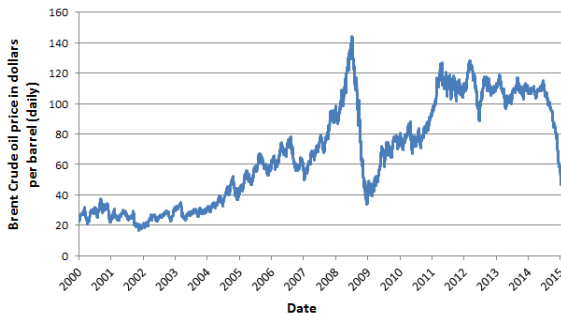
- ¿Cómo comparamos los diferentes turnos de una conversación?
- Comparar uno a uno es un enfoque simplista y no representativo de la realidad

Otro problema que podemos tener es de escalas: por ejemplo, si un hablante es de sexo femenino, su tono será más alto que el de un hombre.

- El método TAMA (Time aligned moving average) ataca estos problemas recién mencionados para la medición de entrainment acústico/prosódico
- ¿Cómo? Construye en primer lugar series de tiempo para cada uno de los hablantes, dada una variable acústico/prosódica.
- Luego aplica herramientas de análisis de series de tiempo para definir alguna medida de entrainment

Series de Tiempo

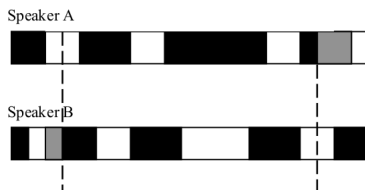
¿Qué es una serie de tiempo?



- En términos coloquiales, una serie de tiempo es una colección de datos temporales.
- Muy frecuentes en Economía y Ciencias de la Atmósfera.
- ¡Mucho más manejables que una sucesión de turnos!

Método TAMA

Cómo construyo la serie de tiempo (dada una variable a-p)

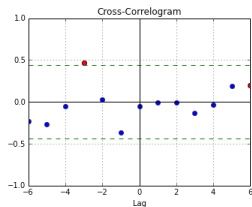
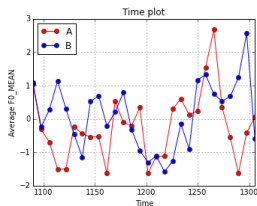


- 1 Partimos la conversación en ventanas solapadas.
- 2 Calculamos un promedio ponderado del valor de la variable acústico-prosódica en cada segmento de habla

$$\mu = \frac{\sum_{i=1}^N f_i d_i}{\sum_{i=1}^N d_i}$$

Método TAMA

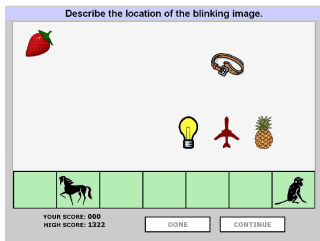
¿Y el entrainment?



- 1 Ya tenemos la serie de tiempo
- 2 ¿Cómo calculamos la mimetización?
- 3 Función de correlación cruzada: mide la influencia de una serie sobre otra.
- 4 Similar a la correlación, pero aplicando un desplazamiento en alguna de las dos series.
- 5 Los valores significativos de esta son los que se consideran los valores de *entrainment* (si es que los hay)

Columbia Games Corpus

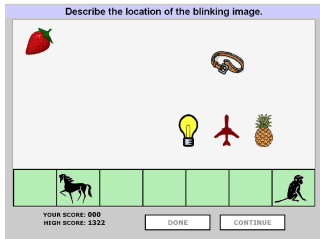
Descripción



- Corpus de conversaciones diádicas en Inglés Americano Estándar
- 12 sesiones con 14 tareas/juegos cada una.
- En cada sesión, se sentó a dos participantes en una cabina profesional de grabación, y una cortina opaca colgando entre ellos para evitar la comunicación visual.
- Los participantes contaron con computadoras a través de las cuales interactuaban mediante juegos.

Columbia Games Corpus

Juegos de objeto



- Dos roles: Descriptor y Seguidor
- En la pantalla, se ven entre 5 y 7 objetos en posiciones aleatorias
- El Descriptor ve uno más, titilante, del cual debe describir su posición
- El Seguidor debe mover la representación del objeto a la misma posición de la pantalla
- Finalmente, se puntúa de 1 a 100 el posicionamiento del objeto

Cinco anotadores escucharon el audio correspondiente a una tarea del juego y respondieron a varias preguntas sobre los sujetos:

Nombre	Pregunta
<i>contributes-to-completion</i>	¿el sujeto contribuye para el éxito del equipo?
<i>engaged-with-game</i>	¿el sujeto parece comprometido con el juego?
<i>making-self-clear</i>	¿el sujeto se expresa correctamente?
<i>planning-what-to-say</i>	¿el sujeto piensa lo que va a decir?
<i>gives-encouragement</i>	¿el sujeto alienta a su compañero?
<i>difficult-for-partner-to-speak</i>	¿el sujeto le hace difícil hablar a su compañero?
<i>bored-with-game</i>	¿el sujeto está aburrido con el juego?
<i>dislikes-partner</i>	¿al sujeto no le agrada su compañero?

De cada una de estas preguntas obtenemos un puntaje de 0 a 5, para cada hablante de cada tarea.

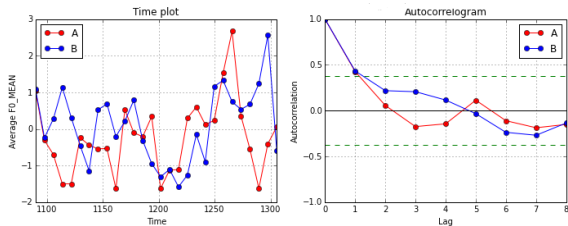
Extracción de features acústico-prosódicas

Usando el software Praat ¹ se extrajeron las variables acústico-prosódicas para cada segmento de habla

Variable	Descripción
<i>F0 Mean</i>	Valor medio de la frecuencia fundamental
<i>F0 Max</i>	Valor máximo de la frecuencia fundamental
<i>Int Mean</i>	Valor medio de la intensidad
<i>Int Max</i>	Valor máximo de la intensidad
<i>NHR</i>	Noise-to-harmonics ratio
<i>Shimmer</i>	Shimmer medido
<i>Jitter</i>	Jitter medido
<i>Sílabas/seg</i>	Cantidad de sílabas por segundo
<i>Fonemas/seg</i>	Cantidad de fonemas por segundo

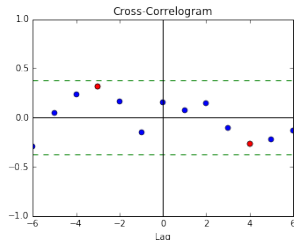
¹<http://www.fon.hum.uva.nl/praat/>

- Usamos un step de 8s y un tamaño de ventana de 16s, manteniendo el solapamiento del 50 %.
- A diferencia del trabajo original de Kousidis, utilizamos series con datos faltantes.
- Pero sólo nos quedamos con aquellas que tengan 5 o más puntos definidos.



Para cada una de las variables acústico/prosódicas

- Calculamos la serie de tiempo para ambos interlocutores por cada tarea
- Calculamos autocorrelogramas, y verificamos su estacionariedad



Definimos dos medidas de entrainment

$$\mathcal{E}_{AB}^{(1)} = r_s \text{ con } s \text{ maximizando } |r_k|, k > 0$$

$$\mathcal{E}_{BA}^{(2)} = |\mathcal{E}_{BA}^{(1)}|$$

Segunda métrica motivada por estudios sobre el disenitainment.