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

Juan Manuel Pérez

22 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 insconsciente 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 insconsciente 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.

Sistemas de diálogo Humano-Computadora

1. Sistemas actuales

Sistemas de diálogo Humano-Computadora

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

Sistemas de diálogo Humano-Computadora

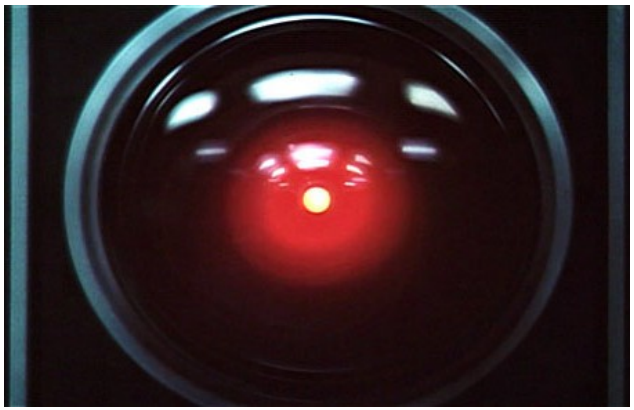
1. Sistemas actuales
2. Bien en la parte lingüística de la comunicación: entender y transmitir mensajes estructuralmente correctos.
3. Mal en la parte superestructural: intercambio de emociones, actitudes, etc.

Sistemas de diálogo Humano-Computadora

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?



Prosodia

- 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

Entrainment

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?

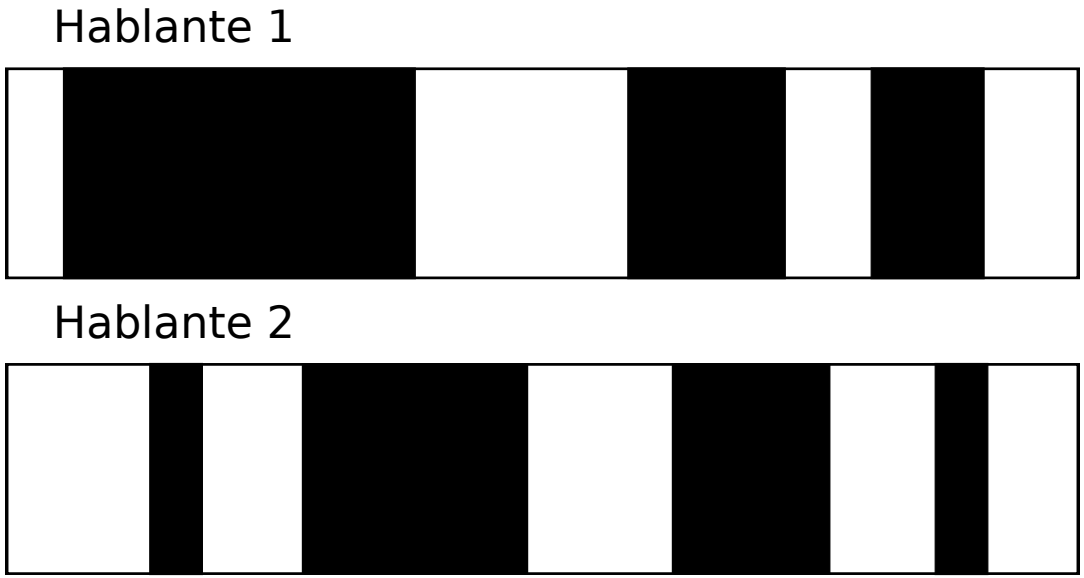
- La definición de entrainment hasta acá vista es muy subjetiva ¿Cómo definimos una medida para esto?
- Vamos a explorar una métrica definida en trabajos anteriores, pulirla un poco, y verificar que efectivamente capture ciertas características del entrainment.
- ¿Cómo? Aplicándola a un corpus con anotaciones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

Problema de la escala

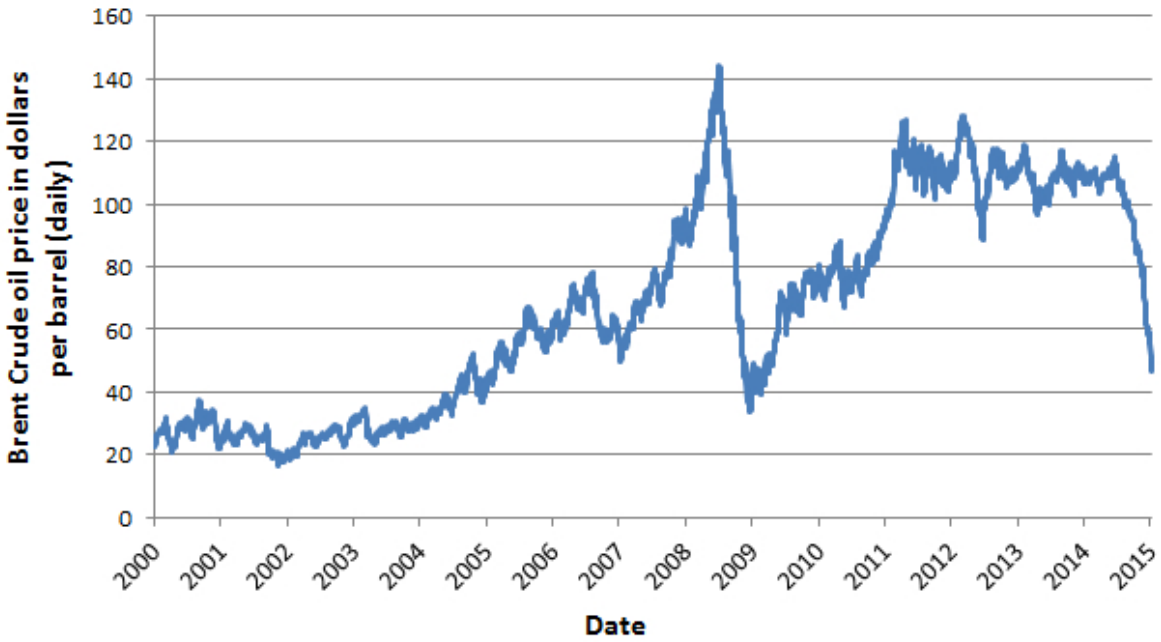
Otro problema que podemos tener es de escalas: por ejemplo, si un hablante es de sexo femenino, su tono será más alto que el de un hombre.

Método TAMA

- 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

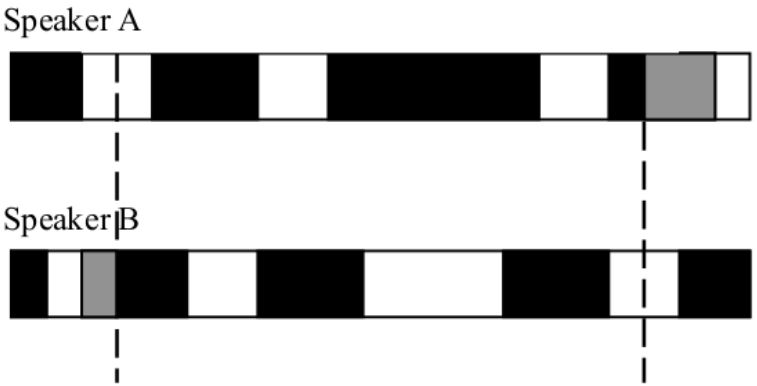
¿Qué es una serie de tiempo?



- 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)

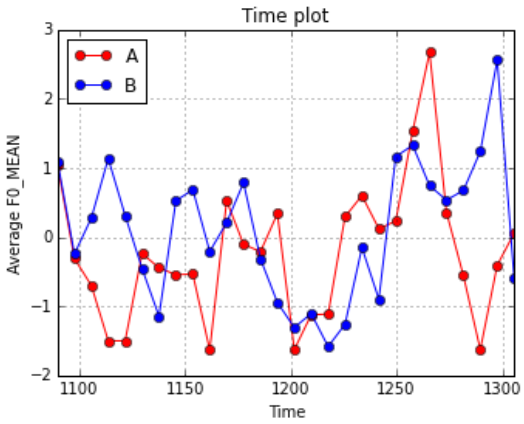


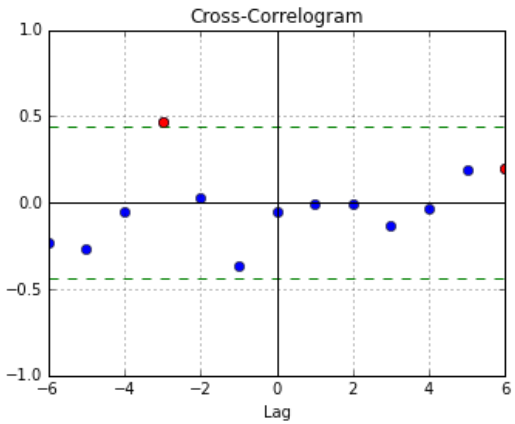
1. Partimos la conversación en ventanas solapadas.
2. Calculamos un promedio ponderado del valor de la variable acústico-prosódica en cada segmento de habla

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

Método TAMA

¿Y el entrainment?

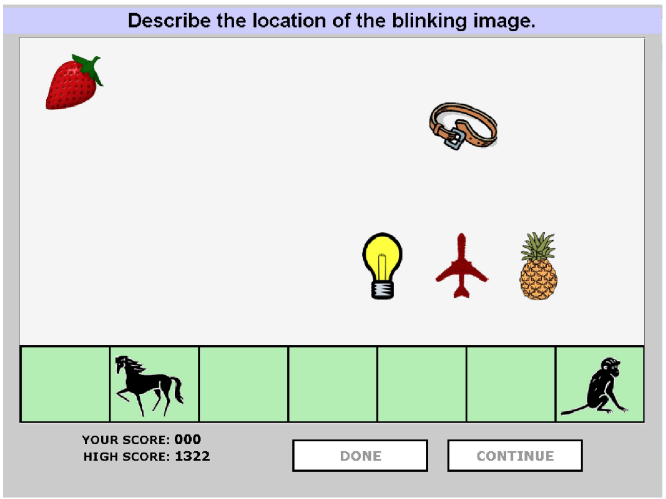




1. Ya tenemos la serie de tiempo
2. ¿Cómo calculamos la mimetización?
3. Función de correlación cruzada: mide la influencia de una serie sobre otra.
4. Similar a la correlación, pero aplicando un desplazamiento en alguna de las dos series.
5. Los valores significativos de esta son los que se consideran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

Columbia Games Corpus

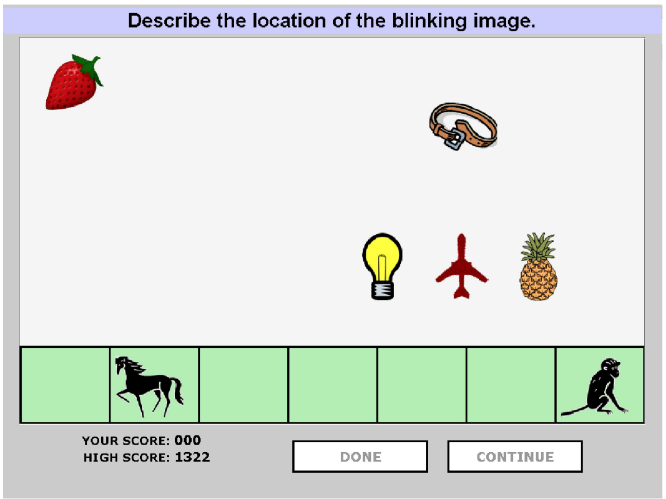
Descripción



- 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

Columbia Games Corpus

Anotaciones sociales

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

Nombre	**Pregunta**
contributes-to-completion	¿el sujeto contribuye para el éxito del equipo?
engaged-with-game	¿el sujeto parece comprometido con el juego?
making-self-clear	¿el sujeto se expresa correctamente?
planning-what-to-say	¿el sujeto piensa lo que va a decir?
gives-encouragement	¿el sujeto alienta a su compañero?
difficult-for-partner-to-speak	¿el sujeto le hace difícil hablar a su compañero?
bored-with-game	¿el sujeto está aburrido con el juego?
dislikes-partner	¿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

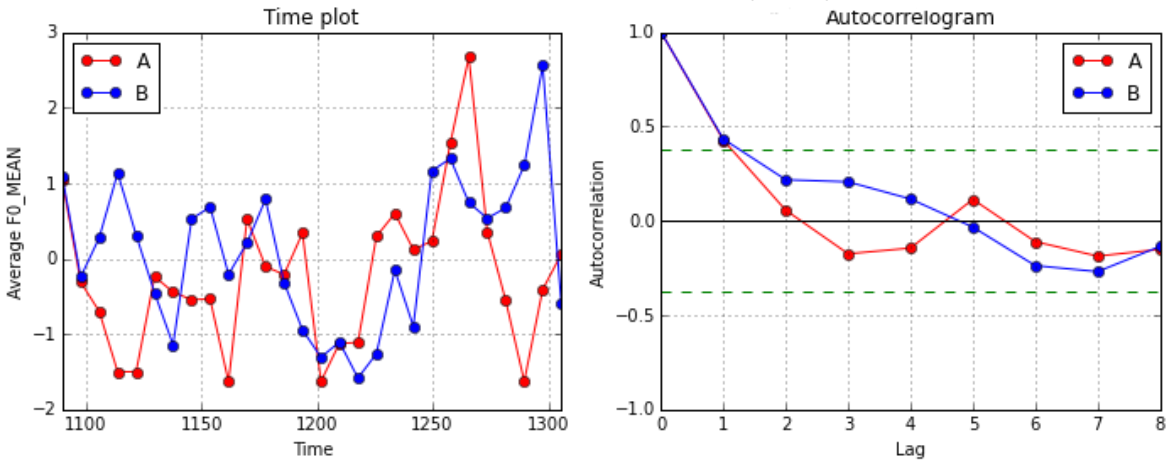
Variable	Descripción
F0 Mean	Valor medio de la frecuencia fundamental
F0 Max	Valor máximo de la frecuencia fundamental
Int Mean	Valor medio de la intensidad
Int Max	Valor máximo de la intensidad
NHR	Noise-to-harmonics ratio
Shimmer	Shimmer medido
Jitter	Jitter medido
Sílabas/seg	Cantidad de sílabas por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

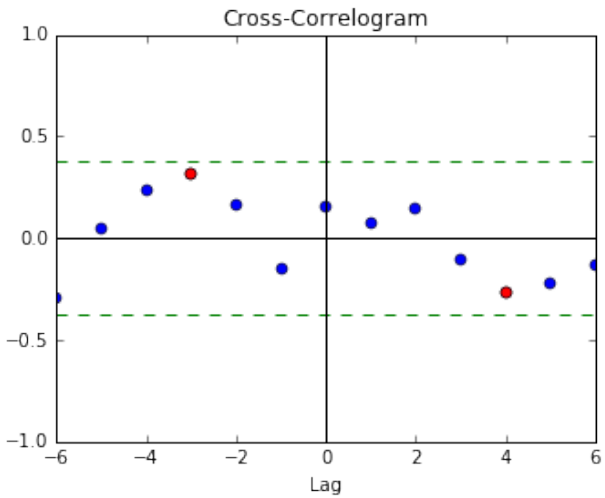
TAMA

Cálculo de series de tiempo



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

$$\mathcal{E}_{AB}^{(1)} = r_s \text{ con s maximizando } |r_k|, k > 0$$
$$\mathcal{E}_{BA}^{(2)} = |\mathcal{E}_{BA}^{(1)}|$$

Segunda métrica motivada por estudios sobre el disentrainment. Healey et al (2014) sugiere que puede ser una conducta de adaptación cooperativa.
Levitan et al (2015) da más indicios en esa dirección.

Entrainment y relación con variables sociales

Para analizar la relación entre las variables sociales (V) y el *entrainment* ($\mathcal{E}$), planteamos un modelo de regresión lineal.

$$V_i \sim \beta_1 + \beta_2 \mathcal{E}_i \tag{1}$$

Nuestra hipótesis es

1. Si SV es una variable de carácter positivo, entonces $\beta_2 > 0$
2. Si SV es una variable de carácter negativo, entonces $\beta_2 < 0$

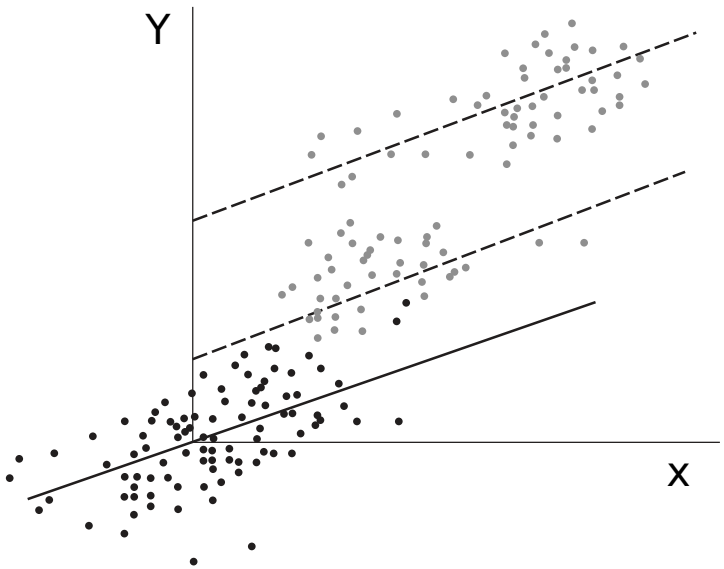
Resultados de Regresión Agrupada

Regresión Agrupada: Hacemos regresión lineal sobre todos los datos, independientemente de su origen.

1. Para $\mathcal{E}_{AB}^{(1)}$ no hubo casi resultados significativos
2. Para $\mathcal{E}_{AB}^{(2)}$ hubo resultados significativos, pero pocos

Para analizar mejor los resultados y eliminar las variables no medidas, efectuamos un análisis de Efectos Fijos.

Regresión Lineal con Efectos Fijos



- Modelo agrupado: niega la posibilidad de heterogeneidad por cada sujeto
- Modelo efectos fijos: heterogeneidad no observada constante en el tiempo para cada sujeto.
- Varias formas equivalentes de calcularlo: Dummy Variable, Within Group.
- En nuestro caso: un sujeto es un hablante dentro de una sesión particular.