

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

Juan Manuel Pérez

23 de marzo de 2016

Sistemas de diálogo Humano-Computadora

Ejemplos de “falta de naturalidad”

Relativamente bien en lo lingüístico,
pero mal en las emociones, actitudes,
intenciones, etc.

Ejemplos:

- 1 Sistemas telefónicos
- 2 Apple Siri
- 3 Google Now
- 4 Amazon Echo
- 5 Microsoft Cortana
- 6 HAL 9000



- 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, refinar y validar una métrica de la mimetización acústico-prosódica.

1 Introducción

2 Antecedentes

3 Materiales y Métodos

- Corpus
- Modificaciones a TAMA

4 Resultados

5 Conclusiones

Prosodia

¿Qué es?

- El “cómo” decimos las cosas, a diferencia del “qué”
- Parte fundamental de la comunicación oral.
- Características que la definen: acentuación, velocidad, tono, ritmo, volumen.
- Una de las principales fallas en los sistemas humano-computadoras hoy día
- Transmite emociones, actitudes, y también desambigua.

Mimetización

¿Qué es?

- 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

Mimetización

¿Y cómo la medimos?

- La definición de mimetización 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 mimetización.
- ¿Cómo? Aplicándola a un corpus con anotaciones sociales, y verificando la relación entre las percepciones sociales y la métrica del *mimetización*

Para el presente trabajo, sólo nos interesa medir la **mimetización acústico-prosódica** sobre variables como el tono o pitch, volumen, calidad de habla, y otras.

Temario

1 Introducción

2 Antecedentes

3 Materiales y Métodos

- Corpus
- Modificaciones a TAMA

4 Resultados

5 Conclusiones

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.

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



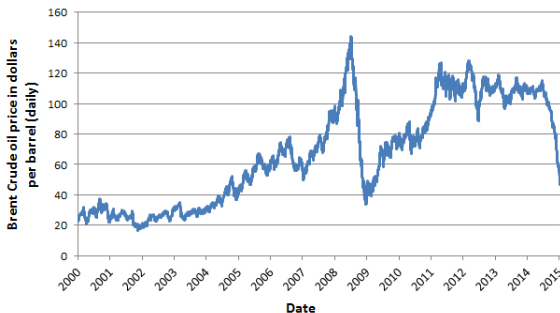
Uno de los problemas que tenemos a la hora de construir métricas de entrainment

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

- Introducido en Kousidis et al (2008).
- 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

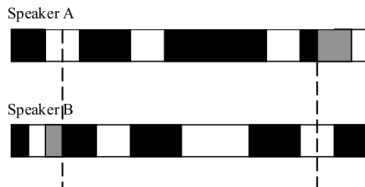
¿Qué es una serie de tiempo?



- 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)

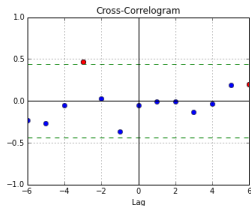
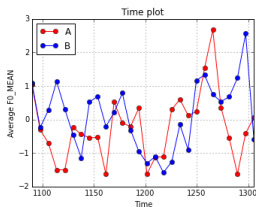


- 1 Partimos la conversación en ventanas solapadas.
- 2 Para cada ventana: calculamos un promedio ponderado del valor de la variable acústico-prosódica en cada segmento de habla

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

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)
- 6 Si el valor es positivo, lo consideramos entrainment
- 7 Si el valor es negativo, lo consideramos disentrainment o antimimetización.

1 Introducción

2 Antecedentes

3 Materiales y Métodos

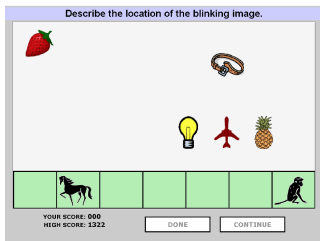
- Corpus
- Modificaciones a TAMA

4 Resultados

5 Conclusiones

Columbia Games Corpus

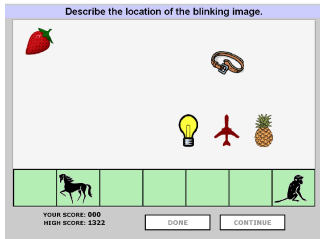
Descripción



- Corpus de conversaciones diádicas en Inglés Norteamericano
- 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.

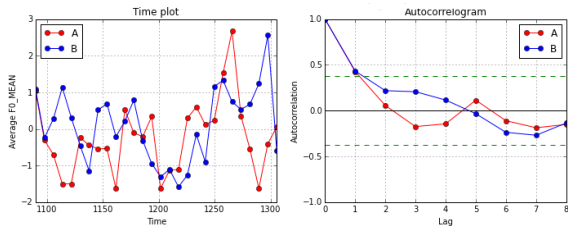
Extracción de features acústico-prosódicas

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/>

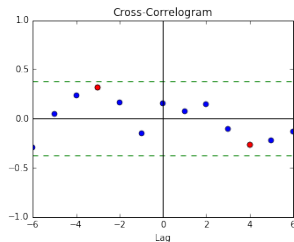
- Usamos un step de 8s y un tamaño de ventana de 16s, manteniendo el solapamiento del 50 %.
- A diferencia del trabajo original de Kousidis, utilizamos series con datos faltantes.
- Pero sólo nos quedamos con aquellas que tengan 5 o más puntos definidos.



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



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

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

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

Segunda métrica motivada por estudios sobre el disenitainment. Healey et al (2014) sugiere que puede ser una conducta de adaptación cooperativa.

Levitan et al (2015) da más indicios en esa dirección.

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

$$V_i \sim \beta_1 + \beta_2 \mathcal{E}_i \quad (1)$$

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 $\beta_2 < 0$

Temario

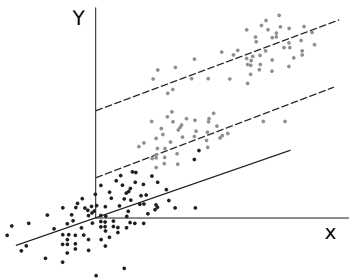
- 1 Introducción
- 2 Antecedentes
- 3 Materiales y Métodos
 - Corpus
 - Modificaciones a TAMA
- 4 Resultados
- 5 Conclusiones

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.

Regresión Lineal con Efectos Fijos

Resultados

	<i>Int Max</i>	<i>Int Mean</i>	<i>F0 Mean</i>	<i>F0 Max</i>	NHR	Fon/seg	Sil/seg	SHIMMER	JITTER
contributes		+	+++	+	++				
clear	+++		+		+++		+		
engaged		+	+++						+
planning									
encourages								+	
difficult	--	++				-			
bored			---		+				
dislikes									

Tabla que resume los resultados significativos del análisis con efectos fijos.
+ indica un valor positivo con significancia $p < 0,10$, ++ con $p < 0,05$ y +++ con $p < 0,01$. Análogamente para los valores negativos.

- 1 Casi todas las variables acústico-prosódicas poseen al menos un valor significativo de $\widehat{\beta}_2$, destacándose *Int Mean*, *NHR* y *F0 Mean*.
- 2 Las variables sociales positivas se relacionan de manera positiva con el valor absoluto del entrainment, en aquellos casos significantes.
- 3 Análogamente ocurre con las variables de connotación negativa (salvo *difficult-for-partner-to-speak* para *Int Mean*)
- 4 La métrica se comporta de manera consistente a medidas de entrainment acústico-prosódico desarrolladas en otros trabajos, como en Gravano et al (2015) utilizando patrones entonacionales.

Temario

- 1 Introducción
- 2 Antecedentes
- 3 Materiales y Métodos
 - Corpus
 - Modificaciones a TAMA
- 4 Resultados
- 5 Conclusiones

- ❶ Desarrollo de métrica de entrainment automática a partir de conversaciones transcritas.
- ❷ Indicios de validación de la métrica introducida por Kousidis et al en un corpus orientado a tareas.
- ❸ Más indicios sobre la prevalencia y característica positiva del disentrainment

- 1 Reproducir experimentos sobre otros corpus, por ejemplo Switchboard²
- 2 Chequear filtro de prewhitening
- 3 Análisis multivariado de las variables acústico-prosódicas para construir nuevas métricas de entrainment

²<https://catalog.ldc.upenn.edu/LDC97S62>