

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

Juan Manuel Pérez Agustín Gravano Ramiro H. Gálvez

Departamento de Computación, FCEyN, UBA

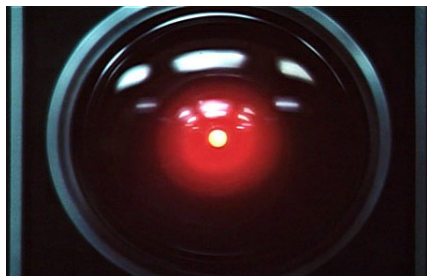
Concurso de Tesis 2016

Sistemas de diálogo Humano-Computadora

Ejemplos de “falta de naturalidad”

Relativamente bien en lo lingüístico,
pero mal en las emociones, actitudes,
intenciones, etc. “Poco naturales”

- 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.

Prosodia

¿Qué es?

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

Mimetización

¿Qué es?



- 1 También conocida como entrainment, convergencia, efecto camaleón, etc.
- 2 Adaptación que ocurre entre hablantes a varios niveles: sintáctico, prosódico, en las posturas, etc.
- 3 Fenómeno ubicuo e inconsciente en la comunicación humana
- 4 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.

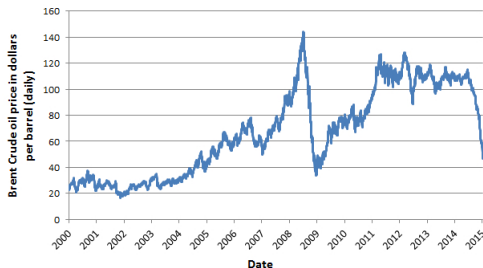
Mimetización

¿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*

Método TAMA

Time Aligned Moving Average

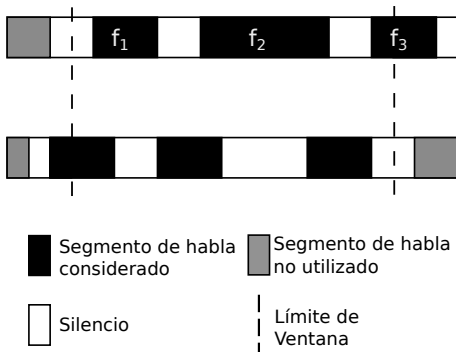


- Introducido en Kousidis et al (2008) ¹
- Construimos en primer lugar series de tiempo para cada uno de los hablantes, dada una variable acústico/prosódica.
- Mimetización se define como una función de estas dos series.

¹ Spyros Kousidis, David Dorrán, Ciaran McDonnell, and Eugene Coyle. Times series analysis of acoustic feature convergence in human dialogues

Método TAMA

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

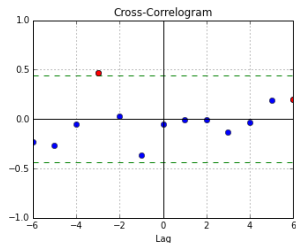
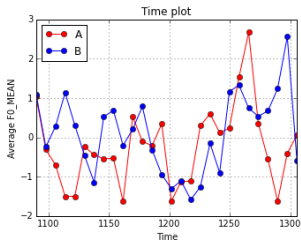


- 1 Partimos la conversación en ventanas solapadas.
- 2 Para cada ventana: calculamos un promedio ponderado del valor de la variable acústico-prosódica en cada segmento de habla

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

Método TAMA

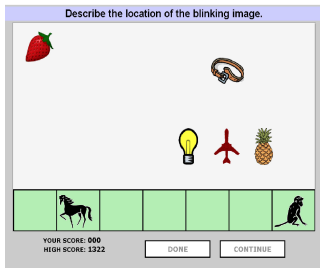
¿Y la mimetización?



- 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 Los valores significativos de esta son los que se consideran los valores de *mimetización* (si es que los hay)

Columbia Games Corpus

Descripción



- Desarrollado por Agustín Gravano para su tesis doctoral.
- 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.

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.

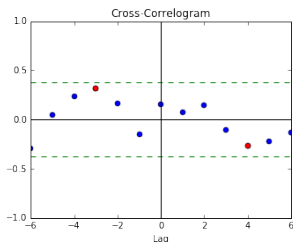
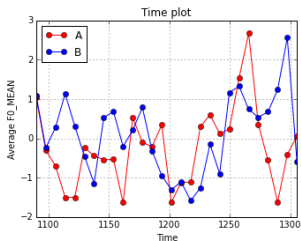
Extracción de features acústico-prosódicas

Usando el software Praat ² 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

²<http://www.fon.hum.uva.nl/praat/>

Definimos dos medidas de mimetización



$$\mathcal{E}_{AB}^{(1)} = \arg \max_{\{r_k, k \leq 0\}} |r_k|$$

$$\mathcal{E}_{BA}^{(1)} = \arg \max_{\{r_k, k \geq 0\}} |r_k|$$

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

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

- Llegado a este punto, ya tenemos medidas de mimetización por cada tarea.
- También tenemos las anotaciones sociales
- ¿Cómo analizamos su relación?

Mimetización y relación con variables sociales

Análisis de Regresión

Para analizar la relación entre las variables sociales (V) y nuestras medidas de *mimetización* (\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 V es una variable de carácter positivo, entonces $\beta_2 > 0$
- 2 Si V es una variable de carácter negativo, entonces $\beta_2 < 0$

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.

Regresión Lineal con Efectos Fijos

Resultados

- ❶ Casi todas las variables acústico-prosódicas poseen al menos un valor significativo de la pendiente.
- ❷ Variables sociales positivas se relacionan de manera positiva con la métrica de mimetización, en aquellos casos significantes.
- ❸ Con variables de connotación negativa ocurre menos claramente.
- ❹ Más indicios sobre el carácter positivo de la antimimetización
- ❺ Comportamiento consistente a medidas de mimetización acústico-prosódicas desarrolladas en otros trabajos, como en Gravano et al (2015) utilizando patrones entonacionales.³

³Agustin Gravano, Stefan Benus, Rivka Levitan, and Julia Hirschberg. Backward mimicry and forward influence in prosodic contour choice in standard american english 2015.

- ❶ Desarrollo de métrica de mimetización 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 antimimetización

- 1 Reproducir experimentos sobre otros conjuntos de datos, por ejemplo Switchboard Corpus. ⁴
- 2 Chequear filtro de prewhitening
- 3 Análisis multivariado de las variables acústico-prosódicas para construir nuevas métricas de mimetización

⁴<https://catalog.ldc.upenn.edu/LDC97S62>

¿Preguntas?