



UNIVERSIDAD DE BUENOS AIRES  
FACULTAD DE CIENCIAS EXACTAS Y NATURALES  
DEPARTAMENTO DE COMPUTACIÓN

# Mimetización entre interlocutores

Tesis presentada para optar al título de  
Licenciado en Ciencias de la Computación

Juan Manuel Pérez

Director: Agustín Gravano

Codirector: Ramiro Gálvez

Buenos Aires, 2015



# MEDICIÓN DE LA MIMETIZACIÓN ENTRE INTERLOCUTORES UTILIZANDO SERIES DE TIEMPO

El *entrainment* (mimetización) es un fenómeno inconsciente que se manifiesta a través de la adaptación de posturas, forma de hablar, gestos faciales y otros comportamientos entre dos o más interactores. A su vez, la ocurrencia de esta mimetización está fuertemente emparentada con el sentimiento de empatía y compenetración entre los participantes.

En esta tesis, nos proponemos explorar una técnica algorítmica para detectar el *entrainment* entre variables prosódicas de dos personas. Esta técnica nos permitirá determinar si existe o no convergencia para ciertos parámetros, y ver como está ésto correlacionado con variables sociales tales como la empatía, la compenetración con la tarea, y otras.

**Palabras claves:** Guerra, Rebelión, Wookie, Jedi, Fuerza, Imperio (no menos de 5).



## MEASURING ENTRAINMENT BETWEEN SPEAKERS USING TIME SERIES

In a galaxy far, far away, a psychopathic emperor and his most trusted servant – a former Jedi Knight known as Darth Vader – are ruling a universe with fear. They have built a horrifying weapon known as the Death Star, a giant battle station capable of annihilating a world in less than a second. When the Death Star’s master plans are captured by the fledgling Rebel Alliance, Vader starts a pursuit of the ship carrying them. A young dissident Senator, Leia Organa, is aboard the ship & puts the plans into a maintenance robot named R2-D2. Although she is captured, the Death Star plans cannot be found, as R2 & his companion, a tall robot named C-3PO, have escaped to the desert world of Tatooine below. Through a series of mishaps, the robots end up in the hands of a farm boy named Luke Skywalker, who lives with his Uncle Owen & Aunt Beru. Owen & Beru are viciously murdered by the Empire’s stormtroopers who are trying to recover the plans, and Luke & the robots meet with former Jedi Knight Obi-Wan Kenobi to try to return the plans to Leia Organa’s home, Alderaan. After contracting a pilot named Han Solo & his Wookiee companion Chewbacca, they escape an Imperial blockade. But when they reach Alderaan’s coordinates, they find it destroyed - by the Death Star. They soon find themselves caught in a tractor beam & pulled into the Death Star. Although they rescue Leia Organa from the Death Star after a series of narrow escapes, Kenobi becomes one with the Force after being killed by his former pupil - Darth Vader. They reach the Alliance’s base on Yavin’s fourth moon, but the Imperials are in hot pursuit with the Death Star, and plan to annihilate the Rebel base. The Rebels must quickly find a way to eliminate the Death Star before it destroys them as it did Alderaan (aprox. 200 palabras).

**Keywords:** War, Rebellion, Wookie, Jedi, The Force, Empire (no menos de 5).



## Índice general

1..	Introducción . . . . .	1
1.1.	Sistemas de diálogo . . . . .	2
1.2.	Mimetización . . . . .	2
1.3.	Midiendo la mimetización . . . . .	3
2..	Series de tiempo . . . . .	5
2.1.	Series de Tiempo . . . . .	6
2.1.1.	Procesos estocásticos . . . . .	7
2.1.2.	Estacionariedad . . . . .	7
2.2.	Descripción TAMA . . . . .	8
2.3.	Análisis Bivariado . . . . .	9
3..	Corpus . . . . .	11
3.1.	Columbia Game Corpus . . . . .	12
3.1.1.	<i>Juego de cartas</i> . . . . .	12
3.1.2.	<i>Juego de objetos</i> . . . . .	13
3.1.3.	Anotaciones sobre comportamiento social . . . . .	13
4..	Método . . . . .	15
4.1.	Selección de Ventana . . . . .	16
4.2.	Time plots . . . . .	18
4.3.	Panel de datos . . . . .	19
4.4.	Análisis de regresión . . . . .	21
4.4.1.	Modelo clásico de Regresión Lineal . . . . .	21
4.4.2.	Nuestro modelo . . . . .	22
5..	Regresión Lineal Agrupada . . . . .	23
5.0.3.	Modelo agrupado o <i>pooled</i> . . . . .	24
6..	Regresión Lineal con Efectos Fijos . . . . .	29
6.1.	Modelo de Efectos Fijos . . . . .	30
6.2.	Definición Formal de Efectos Fijos . . . . .	30
6.3.	Resultados . . . . .	30





## 1. INTRODUCCIÓN

### 1.1. Sistemas de diálogo

Los sistemas de diálogo humano-computadora son cada vez más frecuentes, y sus aplicaciones comprenden una amplia gama de rubros: desde aplicaciones móviles, motores de búsqueda, juegos, o tecnologías de asistencia para ancianos y discapacitados.

Si bien es cierto que estos sistemas logran captar la dimensión lingüística de la comunicación humana, tienen un déficit importante a la hora de procesar y transmitir el aspecto superestructural de la comunicación, que radica en el intercambio de afecto, emociones, actitudes y otras intenciones de los participantes. La habilidad de los participantes de poder expresar, comprender, y reaccionar de acuerdo a estas señales sociales es necesaria para el entendimiento mutuo y una comunicación exitosa.

Un aspecto particular de la comunicación es el fenómeno de *entrainment* (arrastre, mimetización, efecto camaleón), que comprende la adaptación inconsciente de las variables acústicas/prosódicas(a/p) (por ejemplo, el tono de la voz, la velocidad del habla, etc) de manera dinámica en el transcurso de una o varias interacciones. Este fenómeno ha sido introducido por *Brennan et al* [Bre96] en 1996, y se ha observado que la convergencia de los participantes en estas variables ocurre en conjunto con una interacción más fluida y un mayor sentimiento de simpatía por sus interlocutores [CB99].

Poder medir esta mimetización de los interlocutores no es una tarea fácil, sin embargo. En primer lugar, un diálogo no es una sucesión de turnos, sino que es una serie de tiempo dinámica, llena de interrupciones. Más aún, la mimetización no tiene un carácter instantáneo, sino que se sucede a lo largo de la interacción entre los participantes. Estos factores dificultan ostensiblemente poder modelar este fenómeno.

### 1.2. Mimetización

En la literatura de Psicología del Comportamiento se ha observado con frecuencia que, bajo ciertas condiciones, cuando una persona mantiene una conversación, modifica su manera de actuar, aproximándola a la de su interlocutor. En una reseña de este tema se describe a este fenómeno como una “imitación no conciente de posturas, maneras, expresiones faciales y otros comportamientos del compañero interaccional” [CB99, p. 893], y conjeturan que es más fuerte en individuos con empatía disposicional. En otras palabras, personas con predisposición a buscar la aceptación social modifican su comportamiento en forma más marcada para aproximarlo a sus interlocutores.

Esta modificación del comportamiento ha sido observada también en la manera de hablar. Por ejemplo, los interlocutores adoptan las mismas formas léxicas para referirse a las cosas, negociando tácitamente descripciones compartidas, en especial para cosas que resulten poco familiares [Bre96]. Estudios más recientes sugieren que esto también es cierto para el uso de estructuras sintácticas [RKM06]. Este fenómeno subconsciente es conocido como mimetización, alineamiento, adaptación o convergencia, y también con el término inglés *entrainment*, y se ha mostrado que juega un rol importante en la coordinación de diálogos, facilitando tanto la producción como la comprensión del habla en los seres humanos.

### 1.3. Midiendo la mimetización

Muchos estudios han examinado la mimetización del habla, listados en [DLSVC14]. Por ejemplo, [LGW<sup>+</sup>12] propone un método basado en el cálculo de la media de la feature para cada hablante; sin embargo, estos modelos no capturan la esencia dinámica del proceso de *entrainment*.

A la hora de hacer comparaciones razonables entre dos interlocutores, surgen dos problemas [KDW<sup>+</sup>08]. En primer lugar, las curvas tienen diferentes escalas y deben ser normalizadas (por ejemplo, el *pitch* entre un interlocutor masculino y uno femenino), aunque en algunos casos las comparaciones desnormalizadas tienen sentido (volumen). En segundo lugar, surge el problema del alineamiento. ¿Qué partes del diálogo de un interlocutor deberían compararse con qué otras partes? Un approach de comparar interlocuciones uno a uno es demasiado simple y no captura situaciones de diálogo reales.

Para atacar estos inconvenientes, utilizamos el método TAMA (Time Aligned Moving Average), que consiste en separar en ventanas de tiempo el diálogo, y promediar los valores de las variables prosódicas dentro de cada una. Este método es muy similar a aplicar un filtro de Promedio Móvil (Moving Average), lo que da el nombre a la técnica.

Al separar el diálogo en ventanas de tiempo, podemos construir dos series de tiempo en base a cada interlocutor. Estas abstracciones son mucho más tratables que tener una secuencia de elocuciones de parte de cada hablante, y nos permiten efectuar análisis bien conocidos. El *entrainment* podría entonces pensarse, en primera instancia, como la correlación cruzada entre estas series generadas [Cha13].



## 2. SERIES DE TIEMPO

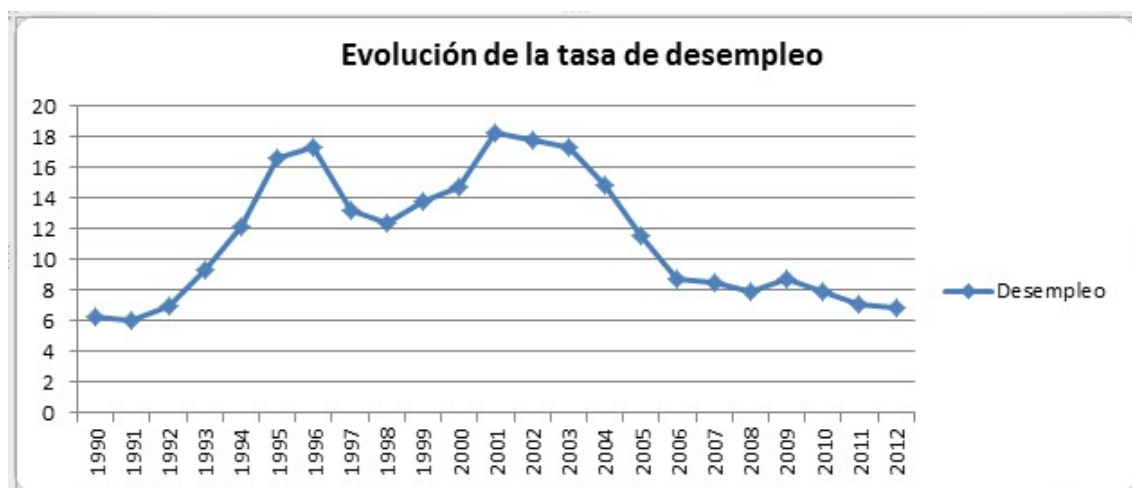


Fig. 2.1: Gráfico de serie de tiempo de la evolución del desempleo en Argentina

## 2.1. Series de Tiempo

### Definición Informal

En términos informales, una serie de tiempo es un conjunto de datos recolectados secuencialmente en el tiempo. Este tipo de datos se dan en varios campos de estudio, mayormente en economía, ciencias de la atmósfera, y otras.

Ejemplos de series de tiempo:

- Volumen de lluvias en sucesivos días de un año
- Precio de acciones en diferentes meses
- Cantidad de habitantes de una ciudad año a año

### ¿Para qué queremos series de tiempo?

Hay varios motivos por los cuales uno querría efectuar un análisis de una serie de tiempo.

1) *Descripción* Usualmente, lo primero que se hace al obtener la serie de tiempo es graficarla y obtener las características más notorias de ésta. Por ejemplo, en 2.1 puede notarse que hay una tendencia decreciente del 2003 hasta el 2012. En otras (como en el volumen de lluvias) podrá observarse cierta estacionalidad en la serie.

Si bien ésto no requiere técnicas avanzadas de análisis, es el primer paso fundamental para comprender una serie de tiempo.

2) *Explicación* Cuando analizamos dos o más series de tiempo, podemos querer ver cómo se comportan en conjunto. Una variación en una serie de tiempo puede producir un cambio en otra. Por ejemplo, podemos intentar buscar como varían en conjunto la temperatura diaria con la cantidad de mL de lluvia caídos.

3) *Predicción* Dada una serie de tiempo, podemos querer intentar predecir un valor futuro.

4) *Control* Dado un proceso del que se mide cierto parámetro de calidad, podemos querer ajustar variables de entrada para mantenerla en ciertos valores.

En nuestro caso, nos es de interés 1 y 2.

### 2.1.1. Procesos estocásticos

**Definición 1.** Un proceso estocástico es una colección de variables aleatorias  $\{X_t\}_{t \in T}$  donde  $T$  es un conjunto de puntos de tiempo. En nuestro caso, nos interesa  $T = \mathbb{N}$ , de manera que el proceso será de la forma  $X_1, X_2, \dots$ .

Podemos entender un proceso estocástico como un conjunto de variables ordenadas por el tiempo. Llamamos serie de tiempo a una observación de este proceso estocástico. Usualmente sólo tendremos esta instancia, a diferencia de otros problemas estadísticos donde tendremos muchas observaciones.

### 2.1.2. Estacionariedad

Un concepto importante en series de tiempo es el de estacionariedad. En lenguaje coloquial, una serie de tiempo estacionaria es aquella en la que no observamos cambios sistemáticos de ésta en el tiempo: si tomamos una parte de la serie, y observamos otro parte distinta de la serie, las propiedades de ésta se mantienen.

Ejemplos de series de tiempo estacionarias son las de ruido blanco, y ejemplos de no estacionarias aquellas que tienen una tendencia. (mejorar esto...)

**Definición 2.** Un proceso estocástico  $X_i, i \in \mathbb{N}$  se dice fuertemente estacionario si, para todo conjunto de índices  $t_1, \dots, t_n$  y para un desplazamiento  $\tau \in \mathbb{N}$  tenemos que

$$F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}} = F_{X_{t_1+\tau}, X_{t_2+\tau}, \dots, X_{t_n+\tau}}$$

Es decir, que la función de probabilidad se preserva por traslados.

Se derivan como propiedades que, para todo  $X_t$  y cualquier desplazamiento  $\tau$

$$E[X_t] = E[X_{t+\tau}] \tag{2.1}$$

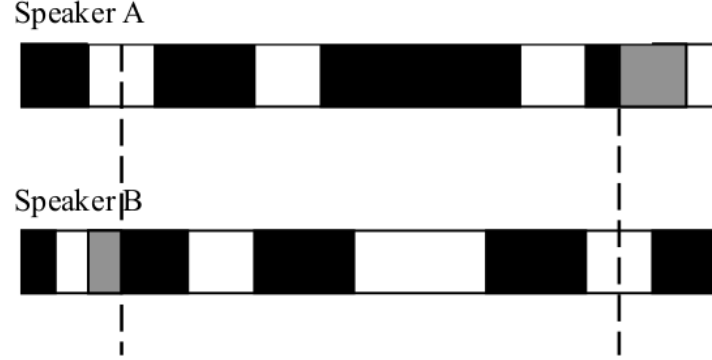
$$Var[X_t] = Var[X_{t+\tau}] \tag{2.2}$$

$$Cov(X_s, X_t) = Cov(X_{s+\tau}, X_{t+\tau}) \tag{2.3}$$

Las ecuaciones 2.1 y 2.2 nos dicen que tanto la media como la varianza son constantes (no dependen de  $t$ ), y que la covarianza sólo depende de la diferencia  $|s - t|$ .

**Definición 3.** Un proceso se dice débilmente estacionario si cumple 2.1, 2.2, 2.3

A partir de aquí, cuando hablemos de series estacionarias estaremos hablando de series débilmente estacionarias



## 2.2. Descripción TAMA

Para construir la serie de tiempo de cada interlocutor, dividimos primero el diálogo en ventanas solapadas de igual tamaño [KDW<sup>+</sup>08]. A la diferencia entre ventana y ventana llamaremos *frame step*, y al tamaño de ventana *frame length*.

Nuestro corpus está anotado de manera que tenemos separadas los intervalos donde los interlocutores hablan (llamaremos a cada uno de éstos locuciones o *utterances*). Para cada una de los frames, calcularemos la media

$$\mu = \sum_{i=1}^N f_i dr_i \quad (2.4)$$

$$dr_i = \frac{d_i}{\sum_{i=1}^N d_i} \quad (2.5)$$

$$(2.6)$$

donde  $i$  itera sobre las locuciones dentro del *frame*,  $d_i$  es la duración de la locución y  $f_i$  es el valor de la *feature* que estamos midiendo.

Como se ve en 2.4, el  $\mu$  que calculamos es una media ponderada por la duración de las locuciones. Así, por ejemplo, al calcular una serie de tiempo sobre el *pitch*, la contribución de interjecciones (usualmente de alto valor) estará disminuída por su breve duración.

La serie de tiempo constará entonces de la secuencia de medias calculadas con 2.4 para cada uno de los frames.



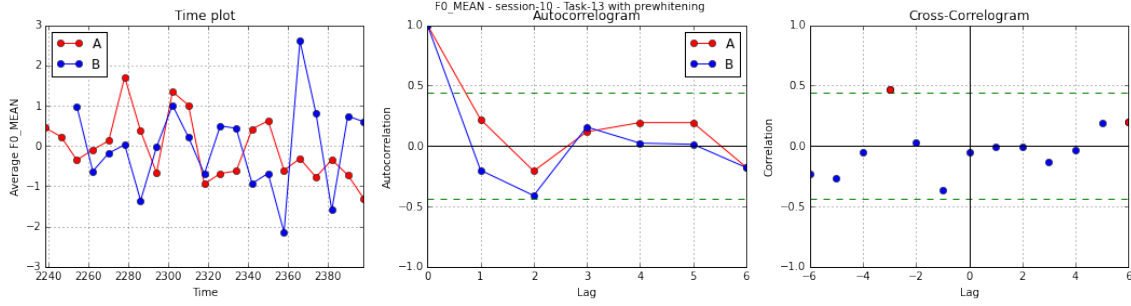


Fig. 2.2: Time-plot producido por TAMA, junto a su autocorrelación y correlación cruzada

### 2.3. Análisis Bivariado

Para medir cuánto se “mimetizan” las dos series, utilizaremos la función de correlación cruzada (c.c.f), que mide cuánto se parecen la serie  $X$  e  $Y$  aplicando un desplazamiento  $k$ , dándonos un valor entre  $-1$  y  $1$  (similar a la correlación de la estadística clásica).

Podemos aproximar la c.c.f. mediante la fórmula de la correlación cruzada muestral.

$$r_{AB}(k) = \begin{cases} \frac{\sum_{t=|k|+1}^n (A_t - \mu_a)(B_{t-|k|} - \mu_B)}{\sqrt{\sum_{t=1}^n (A_t - \mu_a)^2 \sum_{t=1}^n (B_t - \mu_b)^2}} & \text{si } k \geq 0 \\ \frac{\sum_{t=|k|+1}^n (B_t - \mu_b)(A_{t-|k|} - \mu_A)}{\sqrt{\sum_{t=1}^n (B_t - \mu_b)^2 \sum_{t=1}^n (A_t - \mu_a)^2}} & \text{si } k < 0 \end{cases} \quad (2.7)$$

Podemos ver que, si  $k \geq 0$ , lo que hacemos es, a grandes rasgos, calcular la correlación de Pearson entre  $A_{t+k}$  e  $B_t$ . Si  $k < 0$ , lo hacemos entre  $A_t$  e  $B_{t+k}$ .

Para cada tarea, calculamos un correlograma cruzado para  $k \in \{-6, -5, \dots, 0, \dots, 6\}$ . Los valores de  $k \geq 0$  los podemos considerar como aquellos en los cuales nos estamos fijando si  $B$  se mimetiza con  $A$ , y aquellos  $k \leq 0$  al revés. Luego, definimos

$$A \rightarrow B = \max\{r_{AB}(k), k \leq 0\} \quad (2.8)$$

$$B \rightarrow A = \max\{r_{AB}(k), k \geq 0\} \quad (2.9)$$

$$(2.10)$$

$A \rightarrow B$  es el entrainment direccional de  $A$  hacia  $B$ , que mide cuánto se mimetiza  $B$  a  $A$ . (explicar un poco más ésto)



### 3. CORPUS

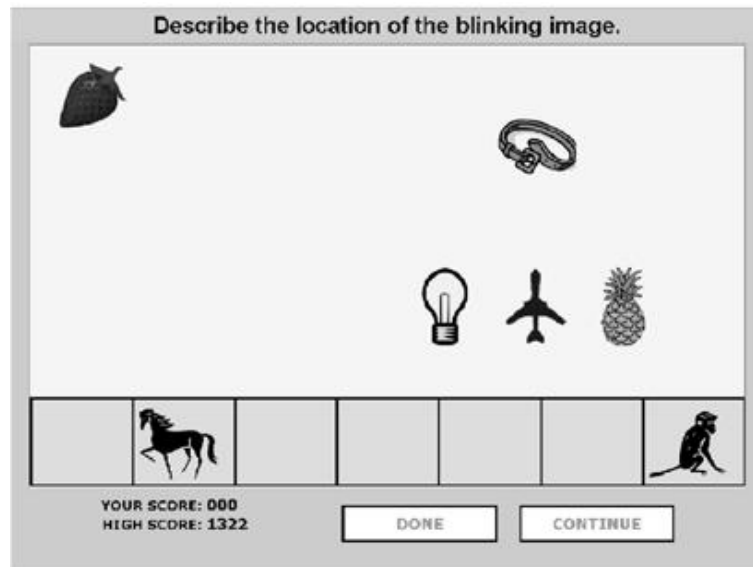


Fig. 3.1: Juego del Columbia Games

### 3.1. Columbia Game Corpus

#### WTF REESCRIBIR ESTO NO TIENE SENTIDO

Nuestro corpus ?? consiste en doce conversaciones diádicas (i.e., con dos participantes) entre trece personas angloparlantes distintas. En cada sesión, se sentó a dos participantes (quienes no se conocían previamente) en una cabina profesional de grabación, cara a cara a ambos lados de una mesa, y con una cortina opaca colgando entre ellos para evitar la comunicación visual.

Las grabaciones se hicieron en 44 kHz, 16 bits con un canal separado para cada hablante; luego fueron guardadas en 16 kHz para el presente estudio. Cada sesión duró aproximadamente 45 minutos, totalizando 9 horas de diálogos, 70.259 palabras (2.037 únicas) para todo el cuerpo de datos.

Los participantes contaron con sendas computadoras portátiles conectadas entre sí, en las cuales jugaron una serie de juegos simples que requerían de comunicación verbal. Por ejemplo, en uno de tales juegos, ambas computadoras muestran un tablero con varios objetos 3.2, todos en la misma posición excepto por uno, el objetivo, que aparece en un lugar distinto en cada computadora.

#### 3.1.1. *Juego de cartas*

En primer lugar, los sujetos del experimento jugaron tres instancias del *Juego de cartas*, donde cada una se subdividía en dos partes. En la etapa inicial del juego, la pantalla de cada jugador mostró un mazo de 9 o 10 cartas. A uno de los jugadores se le pidió que describa la carta en el tope del mazo, mientras que el otro debía buscar en su mazo para encontrar la carta en cuestión, apretando un botón al finalizar ésto. Ésto continúa hasta que las cartas del primer jugador se acaben.

En la segunda parte del *Juego de cartas*, cada jugador tenía una mesa con 12 cartas en ella, inicialmente boca abajo. Mientras el juego comienza, la primera carta de uno de los jugadores se dió vuelta. Este jugador (el Descriptor) debía describir dicha carta al otro

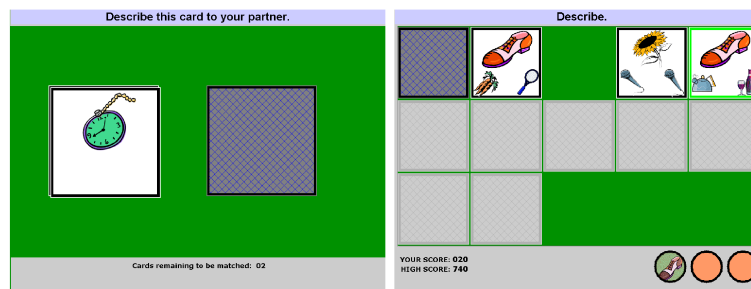


Fig. 3.2: Juego del Columbia Games - Juego de Cartas

jugador (el Buscador), cuya labor era buscar dicha carta en su mesa. Si el Buscador podía encontrar una carta que satisficiera la descripción dada, declararía un acierto y recibiría puntos proporcionales al acierto.

### 3.1.2. *Juego de objetos*

### 3.1.3. **Anotaciones sobre comportamiento social**

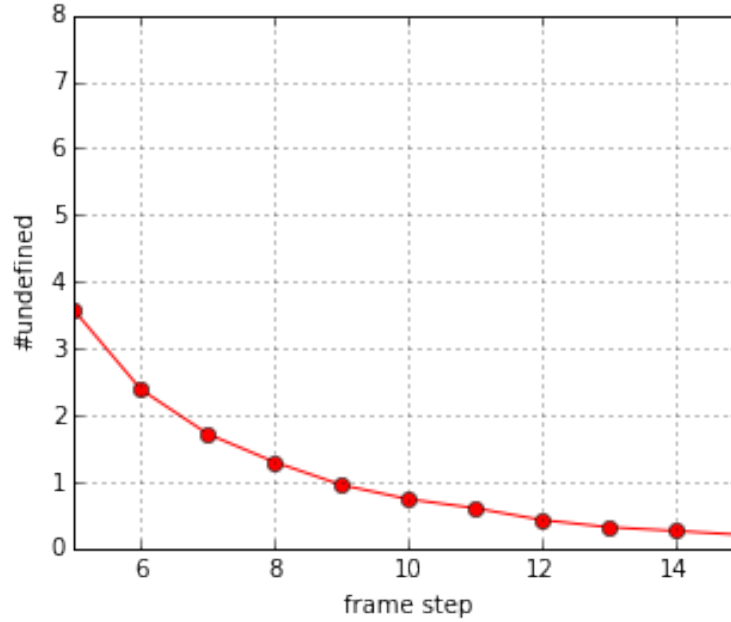
Varios aspectos del comportamiento de los jugadores durante los juegos de Objeto fueron anotados mediante la herramienta de crowdsourcing *Mechanical Turk*. Cada anotador escuchó un clip de un juego y tuvo que responder a las siguientes preguntas (para cada uno de los sujetos):

- ¿el sujeto le hace difícil hablar a su compañero?
- ¿el sujeto parece comprometido con el juego?
- ¿al sujeto no le agrada su compañero?
- ¿el sujeto dirige la conversación?
- ¿el sujeto contribuye para el éxito del equipo?
- ¿el sujeto alienta a su compañero?
- ¿el sujeto se expresa correctamente?
- ¿el sujeto intenta acaparar la conversación?

entre otras. Cada uno de estos clips fue puntuado por cinco anotadores, que respondieron por sí o por no. El puntaje que recibe cada una de las preguntas (a las cuales llamaremos a partir de ahora *variables sociales*) consiste en la cantidad de respuestas afirmativas que recibió, teniendo un rango de 0 a 5.



## 4. MÉTODO



#### 4.1. Selección de Ventana

En [KDMC08] se menciona una elección de *frame step* y *frame length* de 10s y 20s respectivamente. En el caso de nuestro corpus, quisimos buscar los parámetros que mejor se ajustaban a éste, manteniendo la superposición del 50 % entre ventanas sucesivas. Con lo que nos queda que  $FL = 2 * FS$

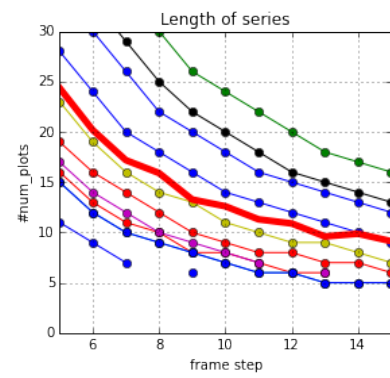
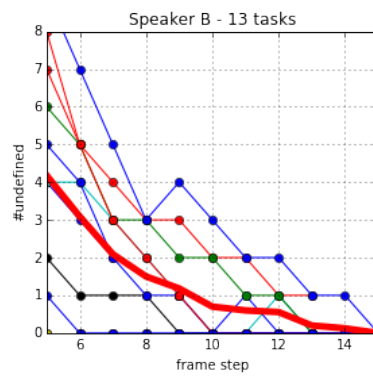
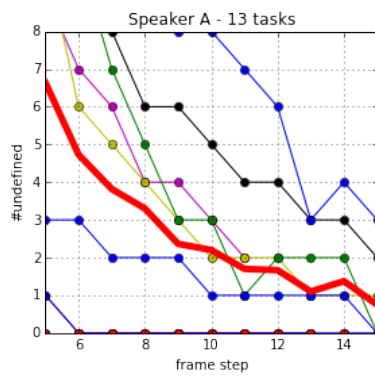
¿Qué queremos optimizar? La métrica que elegimos para esto es encontrar un balance entre un frame no tan grande (para no suavizar en exceso la curva) y que nos reduzca considerablemente la cantidad de indefiniciones; es decir, aquellas ventanas que tomamos en un interlocutor que no tienen ninguna interacción de su parte. Para ver esto, graficamos la cantidad de indefiniciones en función del step tomado.

Dentro del rango de  $FS \in \{5'', 6'', \dots, 15''\}$ , graficamos para cada sesión, tarea y cada interlocutor las curvas de indefiniciones. A su vez, para mayor claridad, graficamos una curva que promedie todas las tareas de una sesión.

Para tener una visión general de lo que ocurría en todas las sesiones, graficamos una curva promedio de todas las sesiones. En ésta puede observarse que hasta  $8'' - 10''$  hay un fuerte descenso de las indefiniciones, que luego se atenúa. Dado que en general tenemos tareas cortas, preferimos tomar  $8''$  como step, y  $16''$  como largo de ventana.

OBS: podríamos cambiar esto a un boxplot!





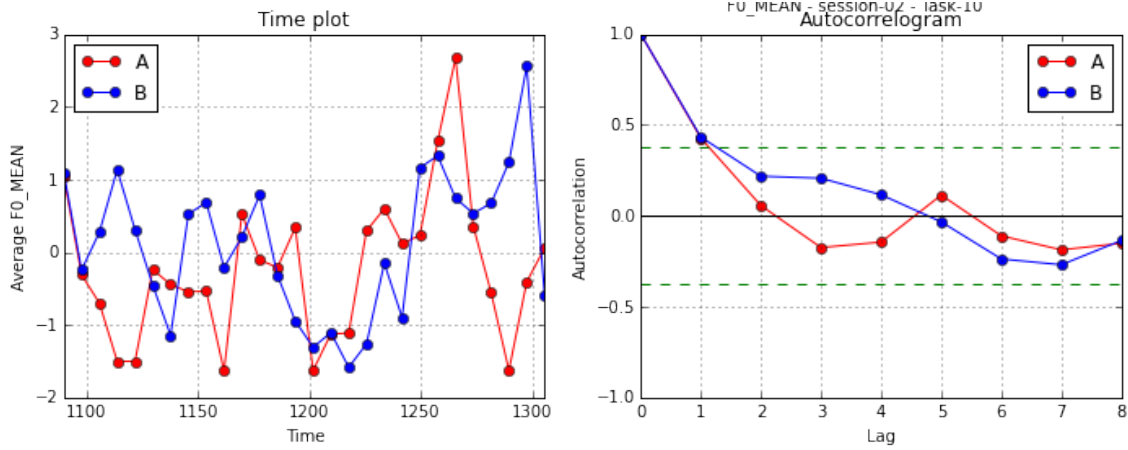


Fig. 4.1: Time-plot producido por TAMA, junto a su autocorrelación

Task	S-01	S-02	S-03	S-04	S-05	S-06	S-07	S-08	S-09	S-10	S-11	S-12
01	—	—	149.888	—	—	—	—	—	54.514	106.096	—	56.135
02	—	—	—	—	—	—	—	—	41.711	63.837	—	—
03	—	51.762	—	80.737	77.977	69.260	68.489	49.607	—	122.272	81.037	—
04	—	187.201	93.333	76.131	79.946	99.240	84.342	—	58.020	129.621	67.977	95.292
05	—	—	—	86.336	—	126.759	145.849	90.742	45.773	134.206	—	—
06	—	—	—	—	—	148.218	50.672	60.281	46.165	66.762	46.773	40.200
07	—	66.024	—	117.762	—	72.410	—	87.702	85.900	110.675	65.758	—
08	—	458.885	98.681	203.867	—	188.708	59.933	48.144	—	157.442	—	81.165
09	—	—	—	75.551	134.247	83.045	108.786	—	62.128	404.014	41.097	92.555
10	50.131	231.392	162.895	242.588	—	122.408	71.198	74.775	—	356.079	69.834	92.769
11	—	74.400	—	98.634	70.189	—	58.911	—	72.947	104.036	59.495	101.970
12	61.331	90.100	129.129	182.917	—	130.375	75.891	57.656	—	101.661	—	64.842
13	55.146	124.095	108.196	144.193	114.720	—	—	83.828	94.087	174.009	84.824	91.525
14	—	75.334	—	—	107.356	—	52.583	144.378	75.589	108.456	91.648	98.487

Fig. 4.2: Tabla de tareas seleccionadas y sus duraciones

## 4.2. Time plots

Usando la técnica descrita, generamos dos series de tiempo para cada tarea. Como antes mencionamos, la ventana elegida es de 16'' con un step de 8'' lo cual da un overlap del 50 %.

Dada una ventana, puede ocurrir que alguno de los interlocutores no haya hablado, o su interacción haya sido demasiado breve como para medir sus variables a/p. En ese caso, y a diferencia de [KDW<sup>+</sup>08], construimos las series sin ese punto, y sin interpolarlo tampoco. (¿por qué no estamos interpolando en vez de dejar los puntos vacíos?)

De estas tareas, sólo nos quedamos con aquellas que tengan al menos 5 puntos definidos para cada serie, de manera que tenga sentido poder calcular la correlación cruzada más adelante (¿podemos justificar un poco más ésto?). Con ésto, no sólo nos interesa la duración de la charla, sino cierta calidad de las series generadas. En 4.2 pueden verse las tareas que tuvimos en consideración, a la vez que su duración.

Los autocorrelogramas de las series bajan rápidamente a cero, un indicio (necesario pero no suficiente) de que las series son estacionarias ¿necesitamos hacer algún chequeo más fuerte de ésto? En [KDMC08] ni se calientan en hacerlo

### 4.3. Panel de datos

Para condensar todos nuestros datos, armamos una tabla por cada variable a/p. Esta tabla contiene información definida para cada interlocutor, tarea y sesión de nuestro corpus.

1. session: número de sesión
2. speaker: 0 si corresponde al interlocutor A; B en otro caso
3. task: número de tarea
4. count: La cantidad de puntos definidos que tiene la serie
5. entrainment: Si  $speaker = 0$ , es  $A \rightarrow B$ ;  $B \rightarrow A$  en otro caso
6. best\_lag: el lag del cross-correlogram donde se logra el *entrainment*
7. tama\_mean: el promedio de la variable

Además, agregamos las variables sociales (relativas al interlocutor) para cada fila:

1. contributes\_to\_successful\_completion
2. making\_self\_clear
3. engaged\_in\_game
4. planning\_what\_to\_say
5. gives\_encouragement
6. difficult\_for\_partner\_to\_speak
7. bored\_with\_game
8. dislikes\_partner

El corpus original cuenta con más variables pero éstas son las únicas que tomaremos en cuenta (citar algo acá!)

En el corpus original, cada variable estaba replicada por cada interlocutor, y por sí o por no, de manera que teníamos:

1. *conversation\_awkward\_A\_yes*
2. *conversation\_awkward\_A\_no*
3. *conversation\_awkward\_B\_yes*
4. *conversation\_awkward\_B\_no*

Ésto nos da una tabla de 210 filas, y 21 columnas. Para cada sesión y speaker, podemos pensar que tenemos una serie de tiempo donde el tiempo es cada tarea, y los datos son el entrainment y las variables sociales. En la jerga econométrica, llamamos a este tipo de datos *de panel*[GP99]: un conjunto de mediciones temporales sobre un mismo sujeto a lo largo del tiempo. En este caso el sujeto es un interlocutor en una sesión, el tiempo son las tareas, y las mediciones son los entrainments

session	speaker	task	entrainment	bored	engaged	encourages	clear
1	0	10	0.581475	0	5	5	5
1	0	12	-0.569677	1	5	5	5
1	0	13	0.533701	2	4	5	4
1	1	10	-0.917101	0	5	2	3
1	1	12	0.467112	0	5	4	2
1	1	13	-0.602364	0	5	4	3
2	0	3	0.520696	0	4	5	5
2	0	4	-0.241060	0	5	4	4
2	0	7	0.743719	0	5	4	5
2	0	8	0.147362	0	5	4	2

Fig. 4.3: Panel de datos

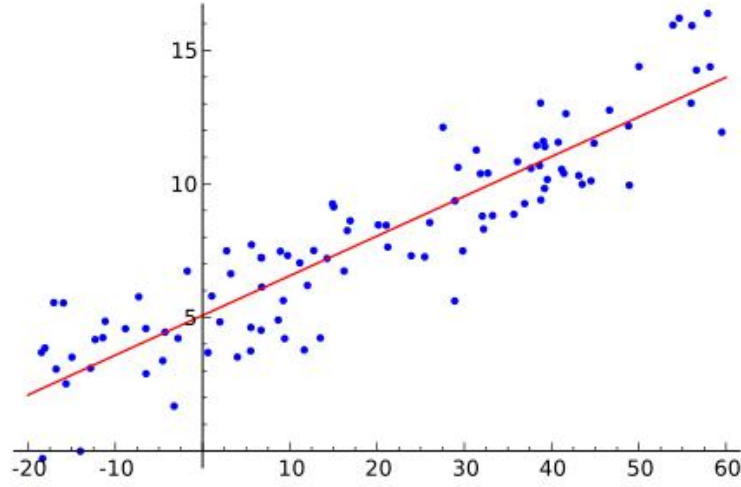


Fig. 4.4: Ejemplo de Regresión Lineal

#### 4.4. Análisis de regresión

Llegado a este punto, dada una variable a/p, nos interesaría evaluar la relación entre el entrainment y las distintas variables sociales. Con esto en mente, planteamos un modelo de regresión lineal donde nuestra variable *explicativa* será la mimetización o entrainment, y la variable *dependiente* será la variable social.

En base a esto, podremos observar cuál es la variación conjunta de ellas. Es esperable que, al aumentar la mimetización (por ejemplo, en la intensidad o en el pitch), aumenten ciertas variables sociales (por ejemplo, la compenetración en el juego) y que otras descendan (el aburrimiento).

##### 4.4.1. Modelo clásico de Regresión Lineal

En el modelo clásico de regresión lineal, tenemos un conjunto de valores fijos  $X_1, X_2, \dots, X_n$ , que son llamadas variables independientes. Asociado a cada uno de estos valores fijos, tenemos variables aleatorias  $Y_1, \dots, Y_n$ . Asumimos, además, que nuestras variables son de la forma

$$Y_i = E[Y|X_i] + u_i \quad (4.1)$$

donde  $u_i$  es la perturbación estocástica de la variable.

Asumiendo que  $E[Y|X_i]$  es una función lineal de  $X_i$ ; es decir, que existen  $\beta_1, \beta_2 \in \mathbb{R}$  que cumplen

$$E[Y|X_i] = \beta_1 + \beta_2 X_i \quad (4.2)$$

obtenemos que

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (4.3)$$

Nuestro objetivo es poder entonces conseguir estimadores  $\hat{\beta}_1, \hat{\beta}_2$  que nos permitan analizar y predecir el comportamiento conjunto de estas variables.

#### 4.4.2. Nuestro modelo

Sea entonces una variable acústica/prosódica (por ejemplo, el pitch o la intensidad), y una variable social de las que acabamos de enumerar en 4.3. Sean  $E_1, \dots, E_n$  los valores de *entrainment* para el set de datos que definimos en 4.3, y sean  $V_1, V_2, \dots, V_n$  los valores de la variable social de cada conversación.

Sobre éstas variables es que planteamos nuestro modelo de regresión lineal clásica: queremos ver qué relación hay tomando como variable “fija” al *entrainment*, y como variable dependiente a la variable social. Queremos hallar, entonces  $\hat{\beta}_1, \hat{\beta}_2 \in \mathbb{R}$

$$V_i \simeq \hat{\beta}_1 + \hat{\beta}_2 E_i \quad (4.4)$$

Para ello, calcularemos los estimadores  $\hat{\beta}_1, \hat{\beta}_2 \in \mathbb{R}$  mediante el método *QR* (insertar referencia aquí) que nos provee el lenguaje R. A su vez, luego de ésto efectuaremos un análisis de significancia sobre  $\beta_2$  para verificar que sean distintos de 0.

Uno esperaría que un alto *entrainment* se relacione con un alto valor de ciertas variables sociales [Bre96], por ejemplo la compenetración con el juego, el ayudar a terminarlo. Esto significa esperar que el valor de  $\hat{\beta}_2$ ; y se relacione con bajos valores de otras, como el aburrimiento, o el rechazo percibido hacia el compañero.

## 5. REGRESIÓN LINEAL AGRUPADA

### 5.0.3. Modelo agrupado o *pooled*

En el modelo agrupado o *pooled*, no distinguimos entre datos provenientes de distintos “grupos” [GP99] y sobre éstos calculamos la regresión lineal, agrupando todos los datos disponibles.

Un problema que surge con este tipo de regresión es que niega todo tipo de *heterogeneidad* de los datos: estos pueden provenir de interlocutores más o menos empáticos, o cuya interacción en el juego se vio influenciada por factores no medidos en el experimento. Todo esto es descartado, aún cuando puede afectar seriamente el resultado obtenido.

AGREGAR GRAFICO DE EJEMPLO PARA ESTO

En el modelo clásico dio resultados con baja significancia. En 5.1 puede verse el gráfico de *F0\_MEAN* y 4 variables sociales y en 5.2 pueden verse los valores de las estimaciones

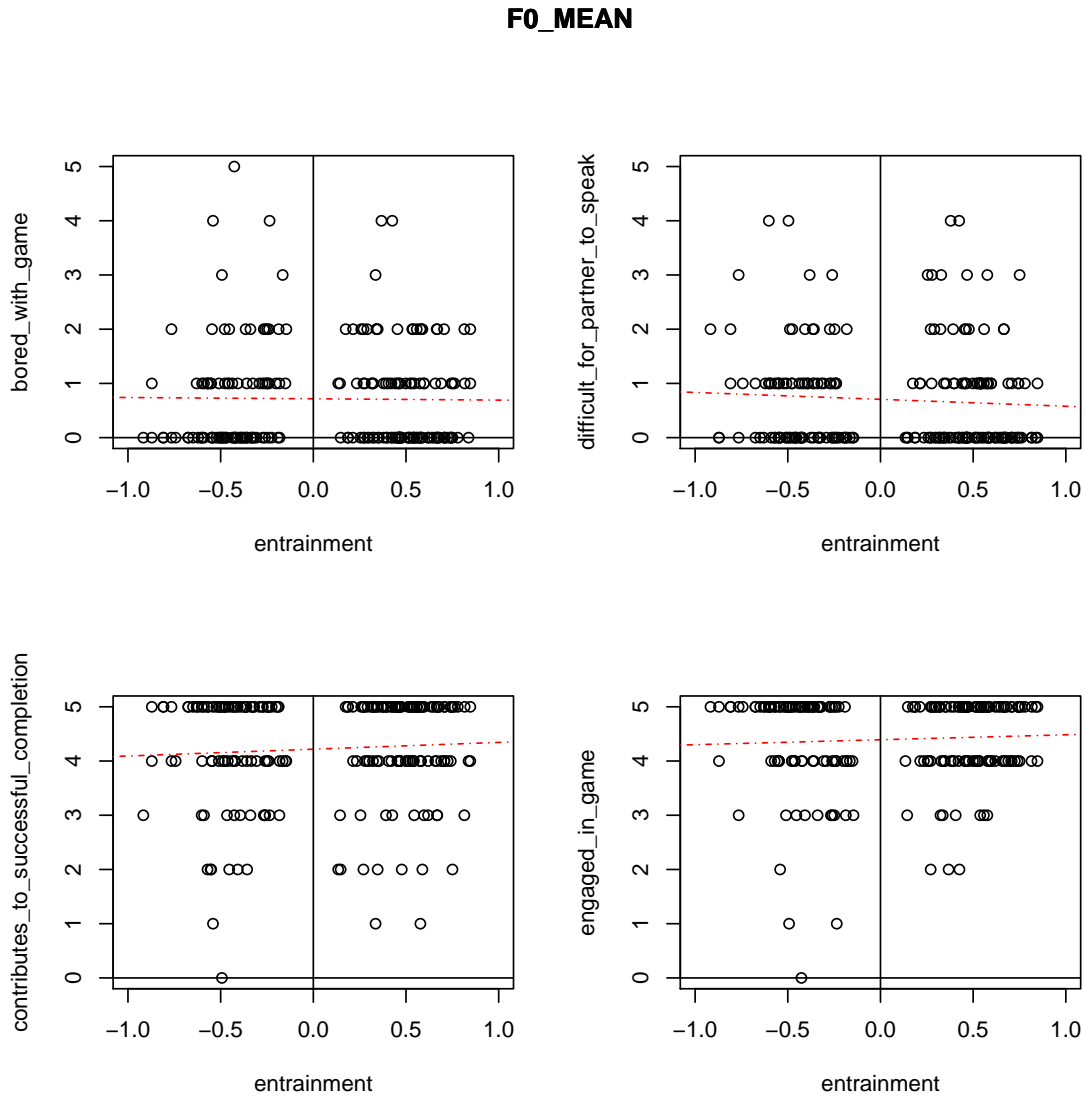


Fig. 5.1: Gráfico de los pares *entrainment*-variable a/p, junto a la regresión lineal obtenida para *F0\_MEAN*



de  $\hat{\beta}_2$  junto a sus p-valores.

ENG_MAX	$\widehat{\beta}_2$	Std. Error	t value	Pr(> t )
bored_with_game	0.0158	0.1327	1.190944E-01	0.9053
difficult_for_partner_to_speak	-0.0053	0.1305	-4.039178E-02	0.9678
contributes_to_successful_completion	-0.0590	0.1401	-4.213426E-01	0.6739
engaged_in_game	0.0618	0.1179	5.240753E-01	0.6008
gives_encouragement	-0.0677	0.1487	-4.552488E-01	0.6494
making_self_clear	0.1466	0.1477	9.927795E-01	0.3220
planning_what_to_say	0.0613	0.1691	3.621580E-01	0.7176
dislikes_partner	-0.1107	0.1129	-9.805928E-01	0.3279
ENG_MEAN	$\widehat{\beta}_2$	Std. Error	t value	Pr(> t )
bored_with_game	-0.0587	0.1326	-4.422931E-01	0.6587
difficult_for_partner_to_speak	-0.0705	0.1304	-5.406454E-01	0.5893
contributes_to_successful_completion	-0.1605	0.1397	-1.149053E+00	0.2519
engaged_in_game	0.0645	0.1179	5.468294E-01	0.5851
gives_encouragement	-0.0064	0.1488	-4.286230E-02	0.9659
making_self_clear	-0.1455	0.1476	-9.858917E-01	0.3253
planning_what_to_say	-0.3036	0.1678	-1.809437E+00	0.0718
dislikes_partner	-0.1061	0.1129	-9.401414E-01	0.3482
F0_MEAN	$\widehat{\beta}_2$	Std. Error	t value	Pr(> t )
bored_with_game	-0.0240	0.1334	-1.801989E-01	0.8572
difficult_for_partner_to_speak	-0.1274	0.1309	-9.732137E-01	0.3316
contributes_to_successful_completion	0.1256	0.1406	8.935204E-01	0.3726
engaged_in_game	0.0911	0.1184	7.694083E-01	0.4425
gives_encouragement	0.1624	0.1492	1.089025E+00	0.2774
making_self_clear	-0.0480	0.1487	-3.228835E-01	0.7471
planning_what_to_say	0.0661	0.1700	3.889381E-01	0.6977
dislikes_partner	-0.1963	0.1129	-1.739084E+00	0.0835
F0_MAX	$\widehat{\beta}_2$	Std. Error	t value	Pr(> t )
bored_with_game	0.3202	0.1302	2.459808E+00	0.0147
difficult_for_partner_to_speak	0.0677	0.1298	5.218418E-01	0.6023
contributes_to_successful_completion	-0.1359	0.1391	-9.770488E-01	0.3297
engaged_in_game	-0.0666	0.1173	-5.673982E-01	0.5711
gives_encouragement	-0.1538	0.1477	-1.041946E+00	0.2986
making_self_clear	-0.3377	0.1454	-2.322852E+00	0.0212
planning_what_to_say	-0.1762	0.1679	-1.049793E+00	0.2950
dislikes_partner	0.1087	0.1123	9.682776E-01	0.3340

Fig. 5.2: Tablas con los resultados de la regresión clásica para ENG\_MEAN, ENG\_MAX, F0\_MEAN y F0\_MAX. En la segunda columna se cita el valor de  $\widehat{\beta}_2$ , la desviación estándar calculada, el t-valor obtenido y la significancia

NOISE_TO_HARMONICS_RATIO	$\hat{\beta}_2$	Std. Error	t value	Pr(> t )
bored_with_game	-0.0575	0.1302	-4.417750E-01	0.6591
difficult_for_partner_to_speak	0.0210	0.1281	1.642415E-01	0.8697
contributes_to_successful_completion	-0.1028	0.1374	-7.487108E-01	0.4549
engaged_in_game	-0.0186	0.1158	-1.608817E-01	0.8723
gives_encouragement	-0.1158	0.1458	-7.942329E-01	0.4280
making_self_clear	-0.0161	0.1453	-1.106847E-01	0.9120
planning_what_to_say	-0.0028	0.1660	-1.687902E-02	0.9865
dislikes_partner	-0.1533	0.1105	-1.387064E+00	0.1669
PHONEMES_AVERAGE	$\hat{\beta}_2$	Std. Error	t value	Pr(> t )
bored_with_game	-0.1237	0.1358	-9.109298E-01	0.3634
difficult_for_partner_to_speak	-0.0188	0.1338	-1.404315E-01	0.8885
contributes_to_successful_completion	-0.0640	0.1437	-4.452462E-01	0.6566
engaged_in_game	0.1098	0.1208	9.094108E-01	0.3642
gives_encouragement	-0.0258	0.1526	-1.690995E-01	0.8659
making_self_clear	-0.0137	0.1518	-9.010196E-02	0.9283
planning_what_to_say	0.1056	0.1733	6.089610E-01	0.5432
dislikes_partner	-0.0409	0.1160	-3.525988E-01	0.7247
PHONEMES_COUNT	$\hat{\beta}_2$	Std. Error	t value	Pr(> t )
bored_with_game	-0.1732	0.1388	-1.247627E+00	0.2136
difficult_for_partner_to_speak	-0.0978	0.1368	-7.151205E-01	0.4753
contributes_to_successful_completion	-0.1309	0.1468	-8.913525E-01	0.3738
engaged_in_game	0.1439	0.1234	1.165972E+00	0.2450
gives_encouragement	0.0320	0.1562	2.051276E-01	0.8377
making_self_clear	0.0730	0.1553	4.703684E-01	0.6386
planning_what_to_say	0.1147	0.1774	6.469016E-01	0.5184
dislikes_partner	-0.0665	0.1187	-5.603928E-01	0.5758
SOUND_VOICED_LOCAL_SHIMMER	$\hat{\beta}_2$	Std. Error	t value	Pr(> t )
bored_with_game	0.0566	0.1324	4.272895E-01	0.6696
difficult_for_partner_to_speak	0.0189	0.1303	1.449885E-01	0.8849
contributes_to_successful_completion	0.0013	0.1399	9.102363E-03	0.9927
engaged_in_game	-0.0017	0.1178	-1.408581E-02	0.9888
gives_encouragement	-0.1514	0.1482	-1.021649E+00	0.3081
making_self_clear	0.1541	0.1474	1.045576E+00	0.2970
planning_what_to_say	0.0753	0.1688	4.459035E-01	0.6561
dislikes_partner	0.0061	0.1129	5.368616E-02	0.9572

SYLLABES_COUNT	$\hat{\beta}_2$	Std. Error	t value	Pr(> t )
bored_with_game	0.0186	0.1355	1.371504E-01	0.8910
difficult_for_partner_to_speak	0.0890	0.1331	6.683149E-01	0.5047
contributes_to_successful_completion	-0.2002	0.1424	-1.405351E+00	0.1614
engaged_in_game	0.0114	0.1205	9.424916E-02	0.9250
gives_encouragement	-0.0264	0.1519	-1.737136E-01	0.8623
making_self_clear	-0.1884	0.1506	-1.251516E+00	0.2122
planning_what_to_say	-0.0688	0.1727	-3.983660E-01	0.6908
dislikes_partner	0.0387	0.1155	3.348662E-01	0.7381
SYLLABES_COUNT	$\hat{\beta}_2$	Std. Error	t value	Pr(> t )
bored_with_game	-0.0320	0.1409	-2.272888E-01	0.8204
difficult_for_partner_to_speak	-0.0693	0.1385	-5.001544E-01	0.6175
contributes_to_successful_completion	-0.1314	0.1485	-8.847251E-01	0.3773
engaged_in_game	0.0468	0.1252	3.739140E-01	0.7088
gives_encouragement	0.0813	0.1578	5.153149E-01	0.6069
making_self_clear	0.1252	0.1569	7.981887E-01	0.4257
planning_what_to_say	0.0095	0.1796	5.275017E-02	0.9580
dislikes_partner	-0.0329	0.1201	-2.743527E-01	0.7841
VCD2TOT_FRAMES	$\hat{\beta}_2$	Std. Error	t value	Pr(> t )
bored_with_game	-0.0224	0.1355	-1.654508E-01	0.8687
difficult_for_partner_to_speak	0.0967	0.1331	7.264700E-01	0.4684
contributes_to_successful_completion	0.0745	0.1430	5.214177E-01	0.6026
engaged_in_game	-0.0432	0.1204	-3.589417E-01	0.7200
gives_encouragement	-0.0955	0.1517	-6.290892E-01	0.5300
making_self_clear	-0.0932	0.1509	-6.175861E-01	0.5375
planning_what_to_say	0.2505	0.1718	1.458374E+00	0.1462
dislikes_partner	-0.0570	0.1154	-4.943188E-01	0.6216

## 6. REGRESIÓN LINEAL CON EFECTOS FIJOS

### 6.1. Modelo de Efectos Fijos

El modelo de efectos fijos agrega el concepto de heterogeneidad permitiendo que cada sujeto tenga su propio valor de ordenada al origen. En el caso concreto de nuestro corpus, dicha heterogeneidad puede deberse a multiplicidad de factores no medidos en él. Por ejemplo, la personalidad de los sujetos es un factor no medido y que puede influir en la dinámica de la interacción entre éstos.

Como el entrainment que medimos mediante el proceso TAMA es un proceso direccional (es decir, medimos tanto la influencia de un interlocutor sobre el otro y viceversa), definimos los “grupos” como las observaciones de entrainment y sus variables sociales de una sesión y el interlocutor sobre el cual consideramos la direccionalidad, tal y cual lo definimos en una sección anterior. Ésto nos arroja la cantidad de 24 grupos de observaciones

### 6.2. Definición Formal de Efectos Fijos

NOTA: Escribir ésto

### 6.3. Resultados

Utilizando como variable explicativa el *entrainment*, los resultados no son significativos. En (NOTA: agregar referencia a las tablas) podemos observar la tabla de coeficientes de esta regresión de efectos fijos.

Por otro lado, este modelo utilizando como variable independiente al valor absoluto del *entrainment* dio valores sustancialmente apreciables. Las variables a/p ENG\_MAX, F0\_MEAN y NOISE\_TO\_HARMONICS\_RATIO poseen valores altamente significativos (p-valor menor a 0.05) para al menos 2 variables sociales. En la tabla 6.1 podemos ver la tabla del test de coeficientes con las variables sociales significativas resaltadas. Una versión simplificada tabla la podemos ver en 6.2 que grafica mediante tabla de doble entrada aquellos pares de variables a/p y variables sociales con coeficientes significativos y su signo.

Con respecto a las variables sociales, podemos observar que:

- *contributes-to-completion* se relaciona positivamente con el *absolute entrainment* cuando la variable a/p medida es F0\_MEAN o bien NOISE\_TO\_HARMONICS\_RATIO. Esto significa que, cuando sube el valor absoluto del *entrainment*, esta variable positiva también lo hace con buena probabilidad. Esto es un efecto esperable: cuando hay mimetización, hay colaboración para el éxito en el juego.
- *making-self-clear*, otra variable que refleja una visión positiva del juego, también se relaciona positivamente con el *absolute entrainment* para las variables F0\_MEAN, NOISE\_TO\_HARMONICS\_RATIO, ENG\_MAX como a su vez para PHONEMES\_AVERAGE y para SYLLABES\_COUNT
- *engaged-with-game*, de la misma manera que las dos anteriores, relaciona positivamente pero sólo con F0\_MEAN
- *planning-what-to-say* y *gives-encouragement*, otras variables positivas, no presentan valores significativos.

ENG_MAX	Estimate	Std. Error	t value	Pr(> t )
contributes_to_successful_completion	0.0497	0.4262	0.1165	0.9074
making_self_clear	1.6581	0.3864	4.2909	0.0001
engaged_in_game	0.3307	0.2576	1.2840	0.2008
planning_what_to_say	0.5005	0.5327	0.9395	0.3487
gives_encouragement	0.4264	0.3792	1.1246	0.2622
difficult_for_partner_to_speak	-0.7200	0.2858	-2.5190	0.0126
bored_with_game	0.2163	0.2560	0.8450	0.3992
dislikes_partner	-0.4318	0.3443	-1.2541	0.2114
F0_MEAN	Estimate	Std. Error	t value	Pr(> t )
contributes_to_successful_completion	1.0274	0.3025	3.3962	0.0008
making_self_clear	0.8307	0.3934	2.1115	0.0361
engaged_in_game	0.8850	0.2750	3.2182	0.0015
planning_what_to_say	0.7167	0.5400	1.3273	0.1860
gives_encouragement	0.0075	0.3941	0.0190	0.9848
difficult_for_partner_to_speak	-0.5975	0.3928	-1.5209	0.1300
bored_with_game	-0.7586	0.2481	-3.0572	0.0026
dislikes_partner	0.0371	0.3800	0.0977	0.9223
NOISE_TO_HARMONICS_RATIO	Estimate	Std. Error	t value	Pr(> t )
contributes_to_successful_completion	0.7041	0.3404	2.0686	0.0400
making_self_clear	1.3344	0.3537	3.7725	0.0002
engaged_in_game	0.0954	0.3462	0.2756	0.7832
planning_what_to_say	-0.1874	0.4177	-0.4485	0.6543
gives_encouragement	0.7234	0.4782	1.5127	0.1321
difficult_for_partner_to_speak	-0.1941	0.3436	-0.5648	0.5729
bored_with_game	0.5876	0.3028	1.9404	0.0539
dislikes_partner	0.3582	0.3330	1.0755	0.2835

Fig. 6.1: Tablas con los resultados de la regresión de efectos fijos para ENG\_MAX, F0\_MEAN y NOISE\_TO\_HARMONICS\_RATIO. En la segunda columna se cita el valor de  $\widehat{\beta}_2$ , la desviación estándar calculada, el t-valor obtenido y la significancia. Las columnas resaltadas corresponden a aquellas significantes

- *difficult-for-partner-to-speak*, una variable que representa una característica negativa de la conversación, se relaciona de igual con el *absolute entrainment* cuando la variable acústico prosódica es ENG\_MAX. Ésto contiene sentido, ya que a mayor mimetización de los interlocutores, la dificultad de éstos para hablar debería disminuir.
- La variable *bored-with-game* se comporta de idéntica manera, sólo que con F0\_MEAN.
- *dislikes-partner* no presenta valores significativos

	ENG_MAX	ENG_MEAN	F0_MEAN	F0_MAX	NOISERATIO
contributes			+		+
clear	+		+		+
engaged			+		
planning					
encourages					
difficult	—				
bored			—		
dislikes					

	PHON_AVG	PHON_COUNT	SHIMMER	SYL_AVG	SYL_COUNT	VCD2TOT
contributes						
clear	+				+	
engaged						
planning						
encourages						
difficult						
bored						
dislikes						

Fig. 6.2: Tabla que representa los resultados significantes del experimento. En una de las entradas, tenemos los nombres abreviados de las variables sociales, y en la otra las variables a/p. El símbolo + representa valor significativo y positivo de la pendiente de la regresión de efectos fijos, mientras que — representa significativo y negativo



## Bibliografía

- [Bre96] Susan E Brennan. Lexical entrainment in spontaneous dialog. *Proceedings of ISSD*, 96:41–44, 1996.
- [CB99] Tanya L Chartrand and John A Bargh. The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology*, 76(6):893, 1999.
- [Cha13] Chris Chatfield. *The analysis of time series: an introduction*. CRC press, 2013.
- [DLSVC14] Céline De Looze, Stefan Scherer, Brian Vaughan, and Nick Campbell. Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication*, 58:11–34, 2014.
- [GP99] Damodar N Gujarati and Dawn C Porter. Essentials of econometrics. 1999.
- [KDMC08] Spyros Kousidis, David Dorran, Ciaran McDonnell, and Eugene Coyle. Times series analysis of acoustic feature convergence in human dialogues. In *Proceedings of Interspeech*, 2008.
- [KDW<sup>+</sup>08] Spyros Kousidis, David Dorran, Yi Wang, Brian Vaughan, Charlie Cullen, Dermot Campbell, Ciaran McDonnell, and Eugene Coyle. Towards measuring continuous acoustic feature convergence in unconstrained spoken dialogues. 2008.
- [LGW<sup>+</sup>12] Rivka Levitan, Agustín Gravano, Laura Willson, Stefan Benus, Julia Hirschberg, and Ani Nenkova. Acoustic-prosodic entrainment and social behavior. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies*, pages 11–19. Association for Computational Linguistics, 2012.
- [RKM06] David Reitter, Frank Keller, and Johanna D Moore. Computational modelling of structural priming in dialogue. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pages 121–124. Association for Computational Linguistics, 2006.