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 esto 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).

ToDo

P.

Índice general

1. INTRODUCCIÓN

1.Robar un poco de Pieraccini y Huerta - Where do we go from here

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.

2.Agregar más ejemplos y referencias

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 oral, que radica en el intercambio de afecto, emociones, actitudes y otras intenciones de los participantes. Éste problema puede verse en cualquier sistema que interactúe sintetizando lenguaje humano: por ejemplo, las aplicaciones telefónicas que atienden automáticamente a sus clientes. Stanley Kubrick y Arthur C. Clarke predijeron esto a la perfección, cuando en "2001: Una Odisea en el Espacio"(1968) dotaron a *HAL* de una voz monótona y robótica, casi lobotomizada. Otro problema grave que sufren estos sistemas humano-computadora es que asumen que sus interacciones de "a turnos", cuando las conversaciones entre humanos suelen distar bastante de ese modelo.

Dentro de las cualidades del lenguaje oral, una de las más distintivas es la *prosodia*, qué es la dimensión que capta *cómo* se dicen las cosas, en contraposición a *qué* se está manifestando. Posee varias componentes acústico/prosódicas: por ejemplo, el tono/pitch, la intensidad o volumen, la calidad de la voz, la velocidad del habla y otras. Un manejo adecuado de éstas componentes es lo que, hoy día, distingue indudablemente una voz humana de una artificial. Ésta carencia de habilidad sobre la prosodia conlleva cierta dificultad en la interacción con agentes conversaciones, que suelen ser calificados como "mecánicos" o "extraños" en su forma de comunicarse.

En pos de mejorar el entendimiento entre agentes conversacionales y sus usuarios, resulta de vital importancia poder entender y modelar las variaciones prosódicas de la comunicación oral. esto se traduciría tanto en una mejor apreciación de lo que quiere comunicar el usuario, como en una mayor naturalidad de la voz sintetizada por el agente.

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, ésta 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 consciente 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

3.Mencionar CAT, fenómeno ubícuo -en teoría-

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. En nuestro caso, nos interesa principalmente el *entrainment* de la prosodia.

1.3. Midiendo la mimetización

Muchos estudios han examinado la mimetización prosódica, listados en [DLSVC14]. Un número importante de ellos se han basado en la premisa de la mimetización como un fenómeno lineal, en el cual la convergencia "va sucediendo" a lo largo de la conversación [BSD95]. Estos estudios dividen las conversaciones en varias partes, y verifican que la diferencia absoluta entre los valores medios (de las variables a/p) y sus desviaciones se aproxime en las últimas partes de la interacción. Sin embargo, este enfoque de la mimetización niega su faceta dinámica: los interlocutores pueden estar inactivos y luego hablar, pueden pasar por varias etapas como escuchar, pensar, discutir un punto, etc. En [LH11] se reportó que éste es un fenómeno no sólamente lineal, sino también dinámico, donde los interlocutores van coincidiendo en el análisis por turnos.

Un problema común que surge a la hora de calcular estas métricas es el hecho de que las conversaciones no están alineadas en el tiempo, ni se dan en turnos. Nos preguntamos entonces qué partes del diálogo de un interlocutor deberían compararse con qué otras partes. Un enfoque de comparar interlocuciones uno a uno es demasiado simple y no captura situaciones de diálogo reales, mucho más dinámicas y con solapamiento casi constante.

Para atacar estos inconvenientes, utilizamos el método *TAMA* (Time Aligned Moving Average) [KDW$^+$08], 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, uno de los cuáles nos permite construir una medida del *entrainment*.

1.4. Objetivo del estudio

En el presente estudio, utilizaremos la técnica de *TAMA* sobre el *Columbia Games Corpus*. Este conjunto de conversaciones consta de doce conversaciones entre dos participantes angloparlantes, quienes interactúan mediante un juego a través de computadoras. El corpus ha sido anotado manualmente con variables que describen la percepción social de la conversación; por ejemplo: ¿el sujeto parece comprometido con el juego? ¿al sujeto no le agrada su compañero?.

Luego de utilizar la técnica *TAMA* para calcular los *entrainment* correspondientes, veremos si existe alguna relación significante entre el valor medido y las percepciones sociales sobre las conversaciones. Uno esperaría, por la literatura previa, que valores altos de *entrainment* se relacionen con valores altos de las variables sociales positivas. A su vez, compararemos los resultados obtenidos con otras medidas de *entrainment* y validaremos la consistencia de la técnica del estudio con éstas.

2. ANTECEDENTES

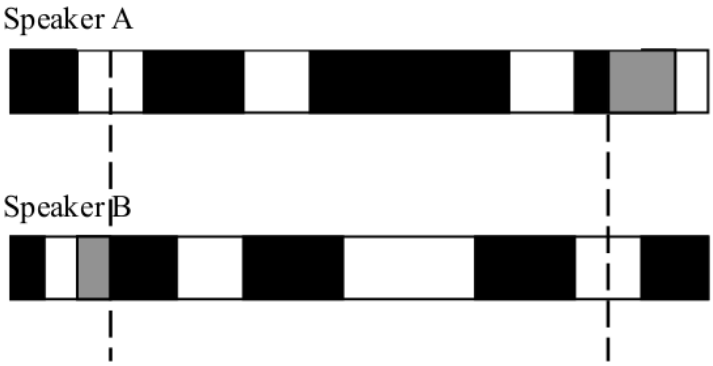

Fig. 2.1: Gráfico de la separación del diálogo en ventanas

En esta sección describiremos el método TAMA desarrollado en [KDW+08] que hemos utilizado como medida de *entrainment* en el presente trabajo.

2.1. Descripción TAMA

En [KDW+08] se introdujo un método novedoso para el análisis del *entrainment* acústico/prosódico. Esta técnica consiste, a grandes rasgos, en armar dos series de tiempo para cada uno de los interlocutores y luego utilizar herramientes de análisis sobre las series construídas. Una serie de tiempo, en términos coloquiales, es una colección cronológica de observaciones, como pueden ser los valores de las acciones de una empresa a lo largo del tiempo, o la cantidad de lluvia medida en *ml* para cada mes de cierto año. En el apéndice 7.1 describimos más en detalle los conceptos básicos sobre series de tiempo.

Un problema que resuelve esta técnica es el del alineamiento: si intentásemos comparar cada segmento del habla (utterance) con otros, ¿cómo los alineamos? Una posibilidad sería uno a uno, aunque esto es muy simplista y poco representativo de la realidad. Al introducir el concepto de series de tiempo, podemos olvidarnos de los segmentos del habla y simplemente utilizar estas construcciones.

Para construir la serie de tiempo de cada interlocutor debemos, en primer lugar, dividir el diálogo en ventanas solapadas de igual tamaño. A la diferencia entre ventana y ventana llamaremos *frame step*, y al tamaño de ventana *frame length*. Consideraremos sólo los segmentos de habla que se encuentren dentro de cada ventana; aquellos segumentos que atraviesen los límites de las ventanas son cortados para que se mantengan dentro de éste. En la figura 2.1 se ilustra el proceso.

Como producto de esto, nuestro corpus queda dividido en una sucesión ventanas solapadas. En el trabajo original, se usa un step de 10 segundos, y un tamaño de ventana de 20 segundos. esto da como resultado un solapamiento del 50 %. En la sección 3.2.1, describimos la elección del tamaño de ventana que hicimos en base al corpus que utilizamos.

Una vez que la conversación se ha partido en ventanas mediante el proceso descripto, se calculan los valores de la serie de tiempo para cada interlocutores de cada una de ellas. esto se hace mediante el siguiente cálculo:

$$\mu = \sum_{i=1}^{N} f_i d'_i \tag{2.1}$$

donde i itera sobre las elocuciones dentro del *frame*, d'_i es la duración relativa del segmento (respecto del tiempo total hablado) y f_i es el valor de la *feature* que estamos midiendo. d'_i se calcula con la fórmula

$$dr_i = \frac{d_i}{\sum_{i=1}^{N} d_i} \tag{2.2}$$

donde d_i es la longitud en segundos de los segmentos del habla en el frame.

Como se ve en 2.1, el valor que calculamos es una media ponderada del valor de la feature por la duración de las locuciones. Así, por ejemplo, al calcular una serie de tiempo sobre la intensidad, la contribución de interjecciones (*ah!* por ejemplo), que suelen tener altos valores *volumen*, estará disminuída por sus breves duraciones.

Una vez obtenidas, dado un feature acústico/prosódico y una conversación, dos series de tiempo mediante el cálculo ventana a ventana de 2.1, necesitamos efectuar algún tipo de análisis sobre éstas para obtener una medida del *entrainment*.

4.Mejorar este dibujo y agregarle una descripción

2.2. Análisis Bivariado

En [KDMC08] se continúa el trabajo en series de tiempo, y se efectúan análisis tanto por cada serie por separado como para las dos en conjunto, lo cual se llama "análisis bivariado" en la terminología de series de tiempo. En este análisis pretendemos analizar ambas series como parte de un sistema, y ver cómo se influyen y retroalimentan mutuamente.

Una posible medida del *entrainment* se podría dar midiendo cuánto influye una serie sobre otra, considerándolas a ambas como parte de un sistema donde ambas interactúan. Éste *entrainment*, entonces, sería direccional: queremos medir cuánto influye el interlocutor A al interlocutor B y viceversa. Puede darse el caso en que ambos tengan fuerte interacción, en tal caso hablamos de *feedback*. En [GBLH15] este concepto es definido como *forward influence*.

5.Agustín, chequea si esto que puse está bien

Para medir cuánto se mimetizan las dos series, utilizaremos la función de correlación cruzada (c.c.f)[Cha13], 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.

$$r_{AB}(k) = \begin{cases} \dfrac{\sum_{t=|k|+1}^{n} (A_t - \mu_A)(B_{t-k} - \mu_B)}{\sqrt{\sum_{t=1}^{n} (A_t - \mu_A)^2 \sum_{t=1}^{n} ([t - \mu_[)^2 B]}} & \text{si } k \geq 0 \\ \dfrac{\sum_{t=|k|+1}^{n} (B_t - \mu_B)(A_{t-k} - \mu_A)}{\sqrt{\sum_{t=1}^{n} (B_t - \mu_B)^2 \sum_{t=1}^{n} (A_t - \mu_A)^2}} & \text{si } k < 0 \end{cases} \tag{2.3}$$

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}. Viéndolo de otra forma,

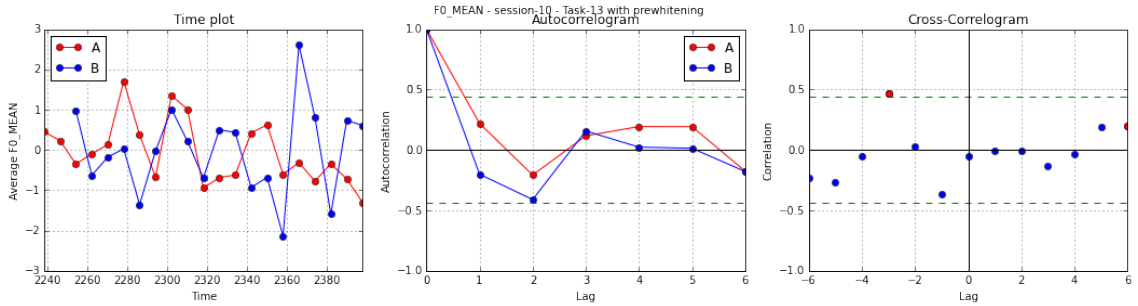

Fig. 2.2: Time-plot producido por TAMA, junto a su autocorrelación y correlación cruzada

si $k \geq 0$, estamos midiendo cuánto influye B sobre A contemplando un desplazamiento de k puntos; si $k \leq 0$ medimos la influencia de A sobre B a misma distancia.

Para cada conversación, se grafica entonces el correlograma cruzado, considerando lags tanto positivos como negativos. Hecho esto, en el estudio [KDMC08] sólo analizan la significancia de los resultados de la correlación cruzada, enumerando aquellos lags en los cuales esto ocurrió. En la sección 3.4 comentaremos cómo utilizamos la técnica descripta para la medición del entrainment direccional.

6.Kousidis menciona algo acerca de los problemas que se dan cuando hay feedback...¿qué deberíamos considerar al respecto de esto?

3. MÉTODO Y MATERIALES

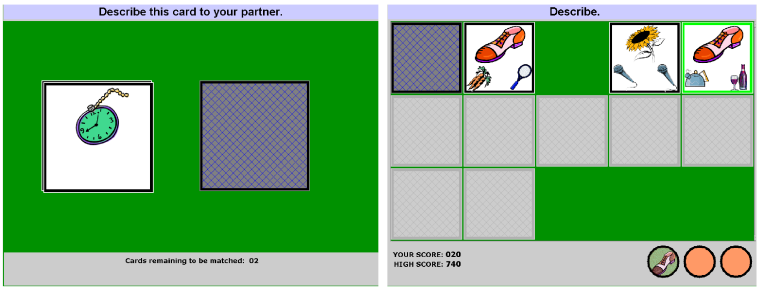

Fig. 3.1: Juego del Columbia Games - Juego de Cartas

En esta sección describiremos tanto el corpus utilizado en el estudio, como así también las modificaciones que efectuamos sobre el método TAMA.

3.1. Columbia Game Corpus

Nuestro corpus [Gra09] consiste en doce conversaciones diádicas (i.e., con dos participantes) entre trece personas angloparlantes distintas. Todos los participantes reportaron hablar Inglés Americano Estándar, y no tener problemas de audición. La edad de los participantes se encuentra en el rango de los 20 a 50 años.

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.

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. Los participantes contaron con sendas computadoras portátiles conectadas entre sí, en las cuales jugaron una serie de juegos simples que requerían de comunicación verbal. El primero de ellos, un juego de cartas que consta de tres instancias que no consideramos en el presente estudio. Luego de esto, pasaron al juego que analizamos.

3.1.1. Juego de Objetos

Luego de completar el juego de cartas, los sujetos siguieron interactuando a través del juego de objetos. En éste, la pantalla de cada jugador mostró un tablero con varios objetos, entre 5 y 7, como se ve en la figura 3.2. Todos ellos se encuentran cercanos, con excepción de uno, a quien llamamos el Objetivo.

Para uno de los jugadores (el Descriptor) el Objetivo apareció en una posición aleatoria entre otros objetos. Para el otro jugador, a quien llamaremos el Seguidor, el Objetivo apareció en la parte baja de la pantalla. Entonces, al Descriptor se le encargó describir la posición del Objetivo de manera que el Seguidor pudiera mover su representación del objeto a la misma posición en su pantalla. Luego de una negociación entre ambos jugadores para decidir la mejor posición del objeto, se les asignó a los jugadores una puntuación entre 1 y 100 puntos de acuerdo a qué tan acertado fue el posicionamiento del Seguidor.

Este juego tomó lugar durante 14 tareas. En las primeras cuatro, uno de los sujetos tomó el papel del Descriptor; en los siguientes cuatro invirtieron roles, y en los finales seis

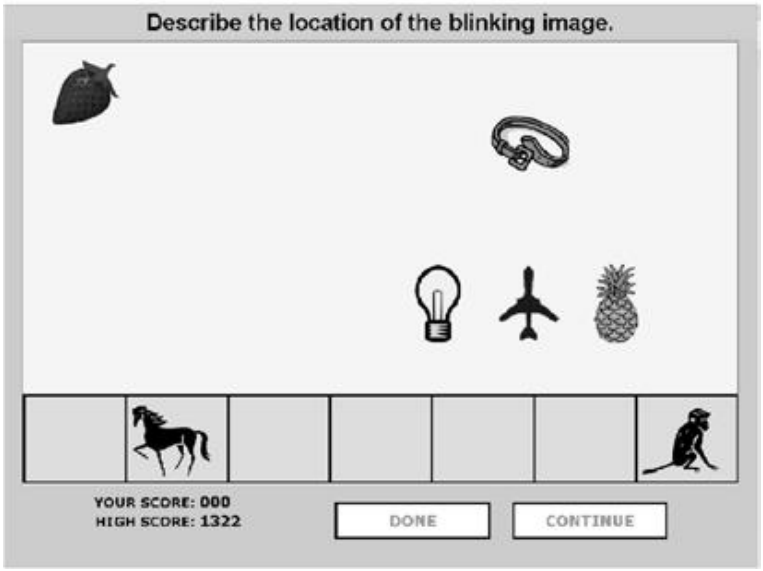

Fig. 3.2: Juego de objetos del Columbia Games

fueron alternando.

3.1.2. Anotaciones sobre comportamiento social

Varios aspectos del comportamiento de los jugadores durante los juegos de objetos 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 parece comprometido con el juego?
- ¿el sujeto dirige la conversación?
- ¿el sujeto contribuye para el éxito del equipo?
- ¿el sujeto alienta a su compañero?
- ¿el sujeto se expresa correctamente?
- ¿al sujeto no le agrada su compañero?
- ¿el sujeto le hace difícil hablar a su compañero?
- ¿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.

3.1.3. Extracción de variables acústico/prosódicas

El set de herramientas *Praat* fue utilizado para extraer automáticamente las variables acústico/prosódicas del corpus, para cada tarea de éste. Las variables que medimos fueron el tono, la intensidad, la proporción de vocalizaciones, jitter, shimmer, cantidad de sílabas, cantidad de fonemas, y la proporción de ruido sobre armónicos. estos atributos fueron medidos en cada uno de los segmentos de habla del corpus.

Repasemos algunos conceptos que necesitamos para definir las variables acústicas.

- *f0* refiere a la frecuencia fundamental de una onda, que es el recíproco del período de ésta. El *tono* o *pitch* es la percepción que tenemos de la frecuencia fundamental, que nos marca cuán agudos o graves son los sonidos.
- *ENG* refiere al volumen o intensidad de la onda. Ésta se mide por la amplitud de la onda, y es la percepción de cuán fuerte es el sonido.
- *jitter* se refiere a los desplazamientos de la onda de la verdadera periodicidad.
- *shimmer* se refiere, de manera similar al jitter, a la variación de la onda pero respecto de la amplitud
- Un *fonema* es la articulación mínima de un sonido, tanto de vocales como de consonantes. Ejemplos de fonemas son los sonidos de las letras u, a, s, k en español.
- El *noise-to-harmonics* ratio puede considerarse como una medida de calidad de la voz, que cuantifica la proporción de ruido que hay en ésta.

En la siguiente tabla describimos más detalladamente cada una de ellas. Recordemos nuevamente que estas features son medidas en un intervalo.

Variable a/p	Descripción
F0_MEAN	Valor medio de la frecuencia fundamental
F0_MAX	Valor máximo de la frecuencia fundamental
ENG_MEAN	Valor medio de la intensidad
ENG_MAX	Valor máximo de la intensidad
VCD2TOT_FRAMES	Proporción de frames vocalizados sobre totales
NOISE_TO_HARMONICS_RATIO	Noise-to-harmonics descripto anteriormente
SOUND_VOICED_LOCAL_SHIMMER	Shimmer medido
SYLLABLES_AVG	Cantidad de sílabas por segundo
PHONEMES_AVG	Cantidad de fonemas por segundo

7.En los lugares donde digo frame, ¿puedo decir fonemas? No explico bien a qué me refiero con frame 8.Cuál es la relación entre NHR y VCD2TOTFRAMES? Uno es 1 -¡el otro¿? 9.Agustín, pasame más tela para cortar de jitter, shimmer y NHR

3.2. Modificaciones a TAMA

Al método TAMA descripto en 2.1 le hemos aplicado algunas variaciones, que pasaremos a detallar.

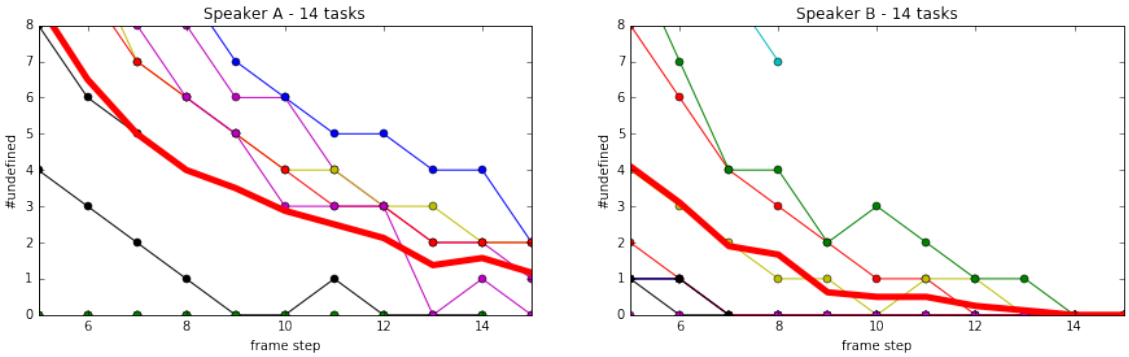

Fig. 3.3: Cantidad de puntos indefinidos en función del step para una sesión en particular, tanto para un interlocutor como para el otro. En rojo se grafica la curva de los promedios.

En primer lugar, [KDMC08] discute la disyuntiva de elegir un tamaño de ventana y step para el método: ventanas demasiado chicas pueden causar que no hayan segmentos de habla en ellas, mientras que un tamaño de ventana demasiado grande suavizaría en exceso la serie de tiempo. A colación de esto, menciona posibles soluciones para el problema de los puntos faltantes: o bien una interpolación (también mencionado en [DLSVC14]) o repetir el punto anterior de la serie.

Estos enfoques, sin embargo, pueden dar lugar a correlaciones artificialmente altas por la construcción misma de la serie. Por otro lado, descartar aquellas conversaciones que tengan puntos faltantes puede ser demasiado restrictivo y eliminar de nuestro corpus una gran cantidad de datos valiosos. Teniendo estas cosas en mente, decidimos aceptar series de tiempo con datos faltantes, que pueden ser producto de ventanas sin segmentos de habla o con algunos demasiado pequeños que imposibilitan la medición de los parámetros (por ej, interjecciones o backchanneling).

10.Agustín, hay alguna traducción de backchanneling?

3.2.1. Selección de Ventana

Otro parámetro de la técnica es el tamaño y step de la ventana, que discutiremos en esta sección. En el trabajo [KDMC08] se menciona una elección de *frame step* y *frame length* de 10s y 20s respectivamente. En el caso de nuestro corpus, queremos buscar los parámetros que mejor se ajustan a éste, manteniendo la superposición del 50 % entre ventanas sucesivas.

¿Qué queremos optimizar? La métrica que elegimos para esto es encontrar un balance entre una ventana 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, para ver qué forma tenían estas curvas. En 3.3 podemos ver las indefiniciones en función de los steps para una sesión. Cada tarea tiene su propia curva, y además graficamos el promedio de todas ellas.

Finalmente, y para tener una visión general de lo que ocurría, graficamos una curva promedio de todas las sesiones, que graficamos en 3.4. Dada ésa curva, intentamos aplicar el "método del codo" para ver si podemos encontrar el valor en el cual la pendiente de las indefiniciones se estanca. Si bien es poco preciso hacer esto, puede observarse que hasta

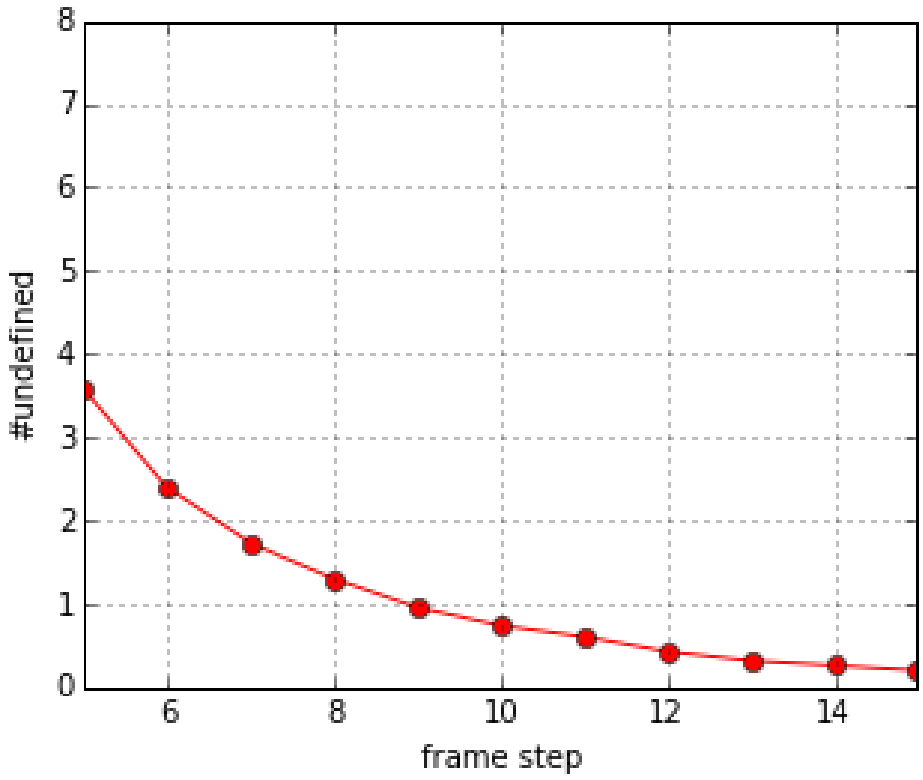

Fig. 3.4: Promedio de cantidad de puntos indefinidos en función del step

$8'' - 10''$ hay un fuerte descenso de las indefiniciones, que luego se atenúa. Dado que en general tenemos tareas cortas, preferimos tomar $8''$ como step, y por ende $16''$ como largo de ventana.

3.3. Time plots

Usando la técnica descripta con las variaciones que consideramos en la anterior sección, 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. Como ya mencionamos en 3.2, y a diferencia de [KDMC08], construimos las series sin ese punto, y sin interpolarlo tampoco.

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. Con esto, no sólo nos interesa la duración de la charla, sino cierta calidad de las series generadas. En 3.5 pueden verse las tareas que tuvimos en consideración, a la vez que su duración.

Como primer paso siempre recomendado en el análisis de series de tiempo [Cha13], graficamos los time plots conjunto de cada par de series, a la vez que sus autocorrelogramas (ver 7.1). En 3.6 podemos observar un ejemplo de esto.

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. 3.5: Tabla de tareas seleccionadas y sus duraciones

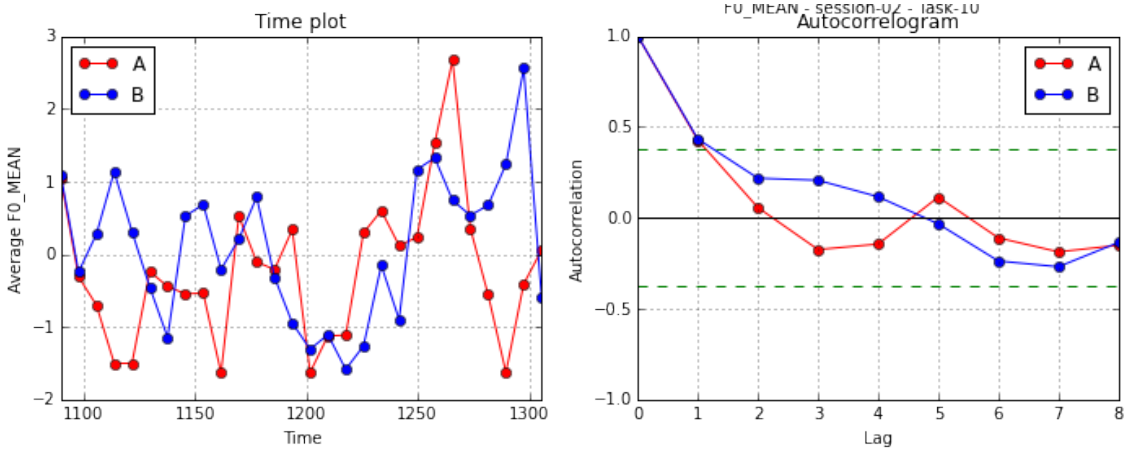

Fig. 3.6: Time-plot producido por TAMA, junto a su autocorrelación

A priori, viendo las series, tienen aspecto de series autoregresivas de orden uno. Es decir, series que son de la forma $X_t = \alpha X_{t-1} + e_t + c$, con e_t ruido blanco, α y c constantes. esto es esperable por la construcción misma del método TAMA, ya que la ventana de cada punto tiene un solapamiento con la ventana anterior. Más aún, uno esperaría que α 0,5 ya que nuestras ventanas tienen ese índice de overlap. Los autocorrelogramas de las series, por otro lado, tienen en su mayoría un valor significativo en $k = 1$, el valor del α de la autoregresión.

El hecho de que los autocorrelogramas desciendan rápidamente a cero es un indicio de que las series de tiempo construídas son estacionarias, lo que nos habilita a efectuar el análisis bivariado de éstas.

3.4. Medición del Entrainment

Considerando todo lo mencionado en 2.2, procedimos a definir una medida de *entrainment* basándonos en el cálculo de la correlación cruzada muestral. Recordemos que, bajo la definición dada en 2.3 de $r_{AB}(k)$, al tomar $k \geq 0$ medíamos cuánto influía B sobre los futuros valores de A, y viceversa cuando $k \leq 0$.

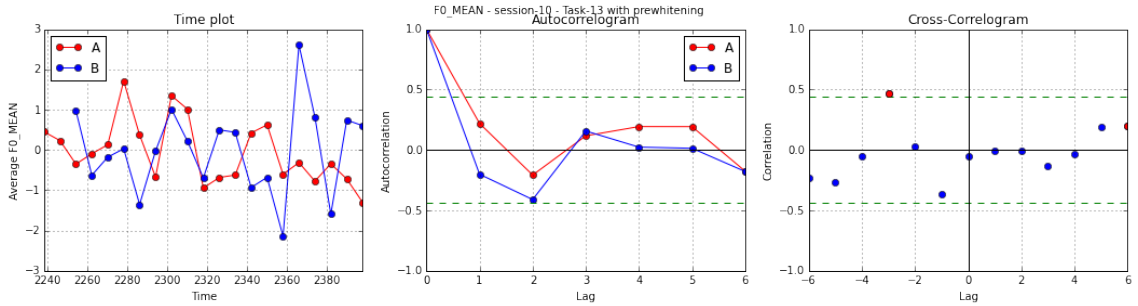

Fig. 3.7: Time-plot producido por TAMA, junto a su autocorrelación y correlación cruzada

Con esto en mente, definimos $\mathcal{E}_{AB}$ como el valor de $r_{AB}(k)$ con mayor valor absoluto, dado $k \leq 0$. Análogamente lo definimos para $\mathcal{E}_{BA}$. En 3.7 podemos observar en rojo los valores de *entrainment* elegidos del correlograma.

Por último, cabe mencionar que a diferencia de [KDMC08] dónde sólo se hacía un análisis de significancia, nosotros vamos a utilizar esta medida independientemente de si es o no estadísticamente diferente de cero.

3.5. Panel de datos

Luego de construir las series de tiempo para cada una de las conversaciones que seleccionamos anteriormente, pasamos a construir una gran tabla que se utilizó en los experimentos de regresión detallados en la siguiente sección. Para condensar todos nuestros datos, armamos una tabla por cada variable acústico-prosódicas, que contiene información definida para cada interlocutor y tarea de nuestro corpus.

Cada columna de esta tabla representa los datos de un interlocutor dentro de una tarea. Éste hecho lo usamos fuertemente a la hora de definir los grupos en nuestro modelo de Efectos Fijos. En la figura 3.5 se describen las columnas generadas.

La tabla generada tuvo una dimensión de 210 x 21, siendo 210 la cantidad de tareas

Campo	*Descripción*
session	número de sesión
speaker	0 si corresponde al interlocutor A; B en otro caso
task	número de tarea
count	La cantidad de puntos definidos que tiene la serie
entrainment	Si *speaker* $= 0$, es $\mathcal{E}_{AB}$; $\mathcal{E}_{BA}$ en otro caso
best_lag	el lag del cross-correlogram donde se logra el *entrainment*
engaged_in_game	¿el sujeto parece comprometido con el juego?
difficult_for_partner_to_speak	¿el sujeto dirige la conversación?
contributes_to_successful_completion	¿el sujeto contribuye para el éxito del equipo?
gives_encouragement	¿el sujeto alienta a su compañero?
making_self_clear	¿el sujeto se expresa correctamente?
planning_what_to_say	
bored_with_game	¿el sujeto se muestra aburrido?
dislikes_partner	¿al sujeto no le agrada su compañero?

Fig. 3.8: Columnas de la tabla generada para ser utilizada en los experimentos de regresión lineal

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. 3.9: Ejemplo de tabla generada para $F0\_MEAN$

(contadas dos veces por cada hablante) y 21 las columnas mencionadas en la figura 3.5. Una forma de ver ésta tabla es que, para cada sesión y hablante, tenemos una serie de tiempo sobre las tareas siendo los datos 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 y las diferentes variables sociales.

En la figura 3.9 tenemos una sección de la tabla. Los sujetos que tenemos en éste ejemplo son 3: *speaker* $= 0$ y *session* $= 1$, *speaker* $= 1$ y *session* $= 1$, y *speaker* $= 0$ y *session* $= 2$. También tenemos cinco series de tiempo para cada sujeto: *entrainment*, *bored*, *engaged*, *encourages* y *clear*. Vale la pena remarcar que estas series de tiempo, al igual que las que consideramos en la construcción de TAMA, pueden tener datos faltantes.

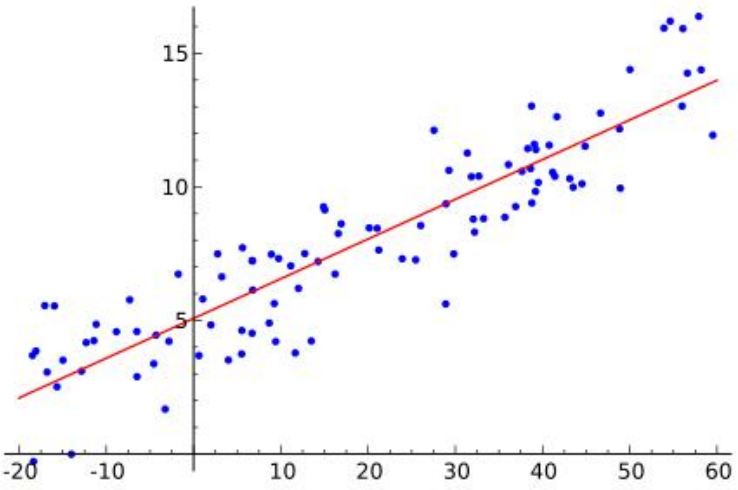

Fig. 3.10: Ejemplo de Regresión Lineal

3.6. Análisis de regresión

Llegado a este punto, dada una variable acústico-prosódica, nos interesó evaluar la relación entre el entrainment y las distintas variables sociales. Con esto en mente, planteamos un modelo de regresión lineal tomando como nuestra variable *explicativa* la mimetización o *entrainment*, y la variable *dependiente* será la variable social elegida. Este análisis de regresión nos permitió observar cuál es la variación conjunta de ellas.

Nuestra hipotesis consistió en que la mimetización (por ejemplo, en la intensidad o pitch) se relacionaría de manera directa con ciertas variables sociales de connotación positiva (por ejemplo, la compenetración en el juego) y que se relacionaría de manera inversa con aquellas de carácter negativo (el aburrimiento o el desagrado por su compañero).

3.6.1. Modelo clásico de Regresión Lineal

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

$$Y_i = E[Y|X_i] + u_i \tag{3.1}$$

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

$$E[Y|X_i] = \beta_1 + \beta_2 X_i \tag{3.2}$$

obtenemos que

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{3.3}$$

Nuestro objetivo es poder entonces conseguir estimadores $\widehat{\beta_1}, \widehat{\beta_2}$ que nos permitan analizar y predecir el comportamiento conjunto de estas variables.

En la siguiente sección se describe el primer experimento, en el cual utilizamos el modelo “pooled” o agrupado, en el cual utilizamos todos los datos juntos indistintamente de la sesión y hablante del que provengan.

4. REGRESIÓN LINEAL AGRUPADA

En esta sección, mostraremos el primer experimento realizado. Este consistió en aplicar un modelo de regresión lineal de cada variable social sobre el *entrainment*, sin desagregar los datos por sesión y hablante.

Una variación que usamos en el presente experimento (y en el posterior) es utilizar como variable dependiente el valor absoluto del *entrainment*, en base a estudios que sugieren que los interlocutores pueden también diferenciarse como un rasgo positivo en la conversación.

4.1. Nuestro modelo de regresión

Dada una variable acústica/prosódica (por ejemplo, el pitch o la intensidad), queremos investigar la relación entre entrainment y las distintas variables sociales medidas. Sea V una variable social de las enumeradas en la tabla 3.5. Sean $E_1, \ldots, E_n$ los valores de entrainment para el set de datos que definimos en la sección 3.5, y sean $V_1, V_2, \ldots V_n$ los valores de la variable social de cada conversación.

Sobre éstas variables es que planteamos nuestro modelo de regresión lineal clásica, para analizar qué relación hay tomando como variable "fija" al *entrainment*, y como variable dependiente a la variable social. El problema, entonces, es hallar estimadores $\widehat{\beta_1}, \widehat{\beta_2} \in \mathbb{R}$ de modo que

$$V_i \simeq \widehat{\beta_1} + \widehat{\beta_2} E_i \tag{4.1}$$

Para ello, calculamos los estimadores mediante el método QR que nos provee el lenguaje R. A su vez, luego de esto efectuamos un análisis de significancia sobre $\widehat{\beta_2}$ para verificar que sea estadísticamente distintos de 0.

Uno esperaría que un alto *entrainment* se relacione con un alto valor de ciertas variables sociales, por ejemplo la compenetración con el juego, el ayudar a terminarlo. En términos de la ecuación 4.1, esperamos que $\widehat{\beta_2} \geq 0$. De manera inversa, cuando las variables sociales tienen un carácter negativo de la conversación, esperamos que $\widehat{\beta_2} \leq 0$.

El modelo de regresión que usamos en este primer experimento se denomina agrupado o *pooled*, ya no distinguimos entre datos provenientes de distintos "grupos" [GP99] y sobre estos 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 influída por factores no medidos en el experimento. Todo esto es descartado, aún cuando puede afectar seriamente el resultado obtenido.

En el siguiente capítulo ahondaremos un poco más en cómo definimos los grupos en nuestro trabajo.

4.2. Resultados sobre *entrainment*

Los resultados de este experimento no fueron interesantes ya que dieron valores de $\widehat{\beta_2}$ con muy baja signifcancia. En 4.1 puede verse el gráfico de regresión lineal del *entrainment* contra distintas variables sociales, tomando como variable acústico-prosódica a F0_MEAN. En las figuras 4.2 y 4.3 podemos observar la tabla de las regresiones, con las estimaciones obtenidas y sus valores de significancia para todas las variables sociales. Se observa que no sólo las pendientes tienen un muy bajo valor absoluto, sino que además ni siquiera tienen los signos que esperábamos en un principio.

En base a lo arrojado por este análisis de regresión, intentamos introducir variaciones del experimento. La primera, es cambiar la variable explicativa por el valor absoluto del entrainment.

11.Mover esto a antecedentes

4.3. Valor absoluto de *entrainment*

En nuestra definición de *entrainment* en el contexto de series de tiempo, la definimos como el valor de la correlación cruzada (en un sentido de los lags) con mayor valor absoluto. Esto puede dar, como resultado, valores positivos entre 0 y 1 a los cuales consideramos como *entrainment*; o bien valores negativos entre -1 y 0, estos considerados como anti-*entrainment*: la divergencia de las features a/p medidas a través del tiempo.

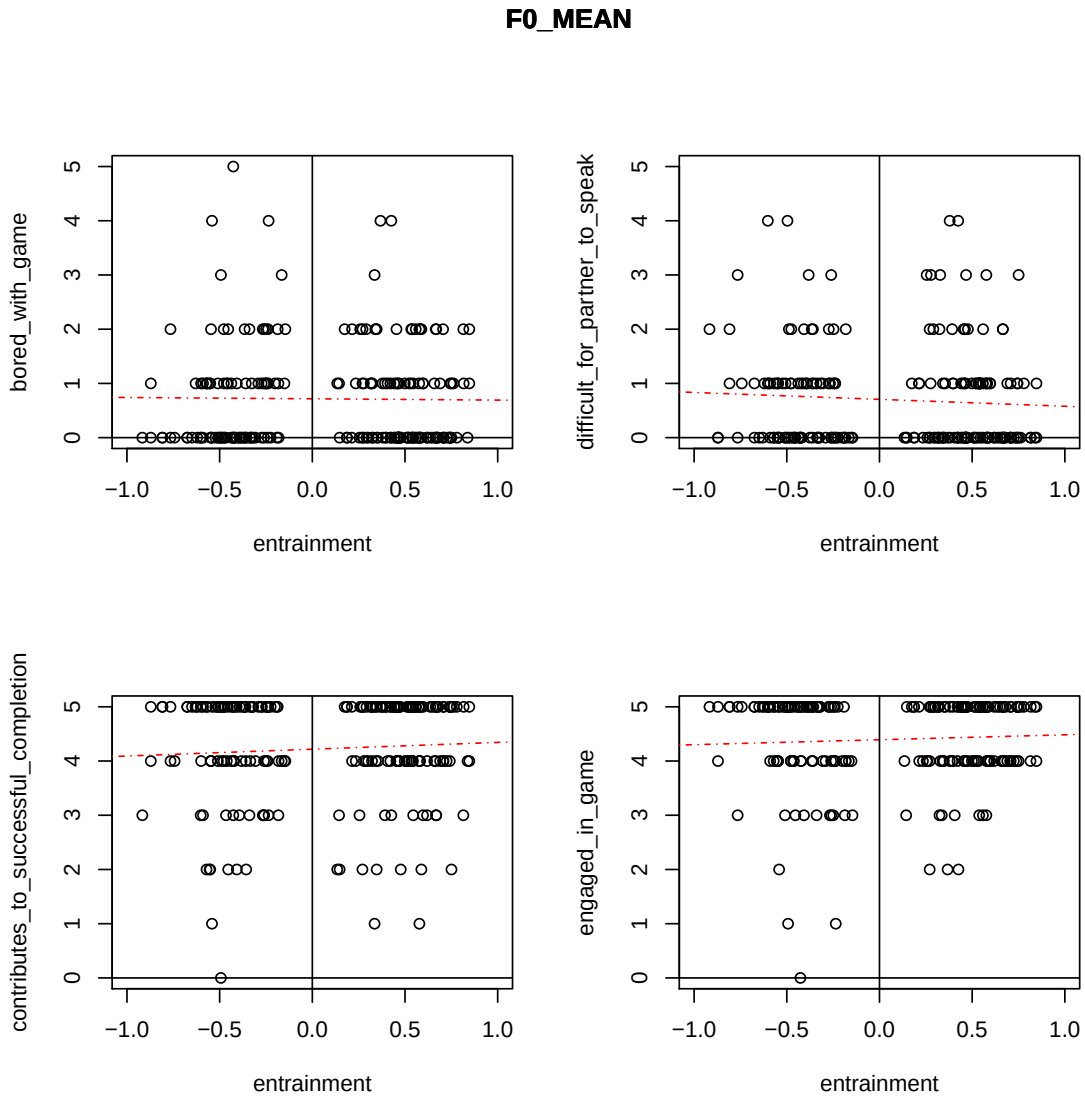

Fig. 4.1: Gráfico de los pares entrainment-variable a/p, junto a la regresión lineal obtenida para *F0_MEAN*

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. 4.2: 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	$\widehat{\beta_2}$	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_AVG	$\widehat{\beta_2}$	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
SOUND_VOICED_LOCAL_SHIMMER	$\widehat{\beta_2}$	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

SYLLABLES_AVG	$\widehat{\beta_2}$	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
VCD2TOT_FRAMES	$\widehat{\beta_2}$	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

Fig. 4.3: 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

Este fenómeno de anti-*entrainment* o antimimicry [CB99] refiere al proceso por el cual uno de los interlocutores no imita al otro sino más bien todo lo contrario, acentúa alguna diferencia. Si bien estudios de larga data como [BGT73] o [DJ69] lo emparentan con una connotación negativa, [HPH14] y [LBGH15] sugieren que puede entenderse este fenómeno como una conducta de adaptación cooperativa. No sólo éso, sino que este fenómeno de mimetización complementaria es más prevalente que la mimetización a secas [LBGH15].

En base esto es que decidimos probar alguna medida que capture positivamente el fenómeno de la anti-mimetización de igual manera que con el *entrainment* antes definido. Es decir, esperamos que cuando tengamos o bien *entrainment* o *entrainment* complementario ocurra que tenemos valores altos de variables sociales de carácter positivo. Mutatis mutandis con las variables sociales de connotación negativa.

Con este fin, en vez de utilizar sólo el valor de *entrainment* como variable explicativa, efectuamos el mismo análisis pero utilizando el valor absoluto del *entrainment* como tal. Usar esto permite captar y valorar el *entrainment* complementario de la misma manera que el "positivo" y valorar su relación con las variables sociales medidas.

4.4. Resultados sobre *absolute value entrainment*

12.Escribir acá, y poner tablas sobre *absolute value entrainment* en pooled

5. REGRESIÓN LINEAL CON EFECTOS FIJOS

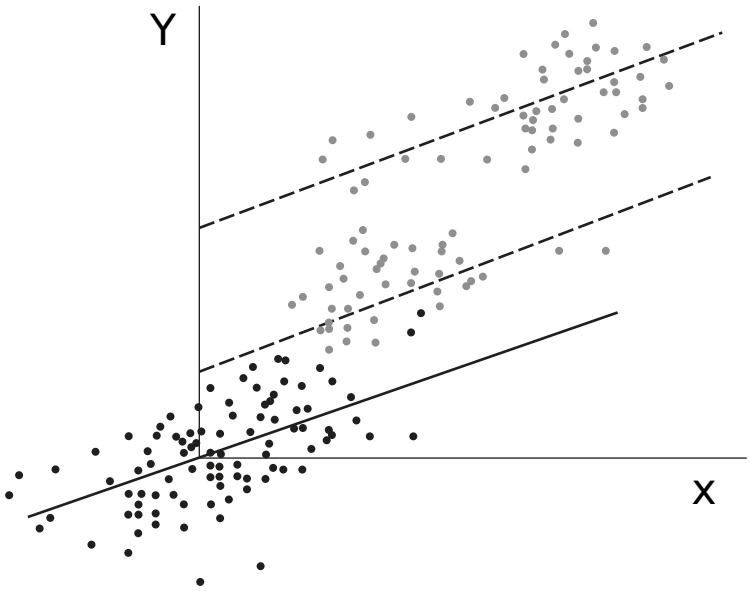

Fig. 5.1: Ejemplo de datos de diferentes sujetos

5.1. Modelo de Efectos Fijos

El modelo agrupado o pooled que vimos en el anterior capítulo niega la posibilidad de heterogeneidad no observada; esto es, variables no medidas que afectan al sistema planteado. En el caso concreto de nuestro corpus, dicha heterogeneidad puede deberse a multiplicidad de factores: la personalidad de los sujetos, el sexo, las elecciones gramaticales, etc.

13.Agustín, chequeá esto por favor

Los modelos de efectos fijos nos ayudan a controlar la heterogeneidad no observada cuando ésta es constante en el tiempo, dado un sujeto del sistema. Asumimos que estos factores son inherentes a la conversación entre el hablante y su par. Para nuestro experimento, definimos los sujetos (en el lenguaje del modelo estadístico) como cada uno de los interlocutores y sus respectivas sesiones. No nos importa si el mismo sujeto se repite en otra sesión: cada hablante de una sesión es un sujeto distinto para el modelo de efectos fijos.

Recordemos que el entrainment medido mediante el proceso TAMA es un proceso direccional: medimos tanto la influencia de un interlocutor sobre el otro y viceversa. Así que el *entrainment* de cada "fila" de nuestro set de datos (definido en la sección 3.5) corresponde al valor de mimetización direccional del actual hablante sobre su par.

5.2. Modelo de efectos fijos *within group*

Existen (al menos) dos variantes del modelo de efectos fijos: el modelo de variables ficticias y el modelo "dentro del grupo" (*within group*). Ambas técnicas son matemática equivalentes, pero utilizaremos la segunda que nos provee el estimador de la pendiente, que es lo único que nos interesa.

La técnica utilizada consiste en, dentro de cada grupo, restarle tanto a la variable "de-

pendiente" y a la "independiente" las medias dentro de cada grupo (de aquí proviene el nombre del método). Ésto resulta en "sacarle" los efectos fijos no-temporales: en nuestro caso, estos están contemplados dentro de la ordenada al origen. Luego de efectuar esta transformación, se aplica regresión lineal "pooled", como en el experimento anterior. Producto del pre-proceso de los datos, la ordenada al origen es negligible. En [GP99, chap 16] se describe extensivamente este procedimiento.

5.3. Resultados

Utilizando como variable explicativa el *entrainment*, los resultados no son son significativos. En **??** podemos observar la tabla de coeficientes de esta regresión de efectos fijos.

14.agregar referencia a tablas de efectos fijos sobre entrainment

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 **??** podemos ver la tabla del test de coeficientes con las variables sociales significativas resaltadas. Una versión simplificada tabla la podemos ver en 5.4 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 value 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 value entrainment* para las variables F0_MEAN, NOISE_TO_HARMONICS_RATIO, ENG_MAX como a su vez para PHONEMES_AVG

- *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.

- *difficult-for-partner-to-speak*, una variable que representa una característica negativa de la conversación, se relaciona de igual con el *absolute value entrainment* cuando la variable acústico prosódica es ENG_MAX. esto contiene sentido, ya que a mayor mimetización de los interlocutores, la dificultad de estos 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

Esto confirma la hipotesis de que el valor absoluto del *entrainment* se relaciona de manera positiva con atributos sociales de características positivas, mientras que lo hace de manera inversa con los que tienen connotaciones negativas.

ENG_MAX	$\widehat{\beta_2}$	Std. Error	t value	Significance
contributes_to_successful_completion	0.0720	0.4258	1.689631E-01	0.8660
making_self_clear	1.6914	0.3820	4.427376E+00	0.0000
engaged_in_game	0.3456	0.2528	1.367266E+00	0.1732
planning_what_to_say	0.5655	0.5208	1.085851E+00	0.2790
gives_encouragement	0.4739	0.3744	1.265523E+00	0.2073
difficult_for_partner_to_speak	-0.6925	0.2863	-2.418510E+00	0.0166
bored_with_game	0.2110	0.2543	8.298495E-01	0.4077
dislikes_partner	-0.4254	0.3438	-1.237312E+00	0.2175
ENG_MEAN	$\widehat{\beta_2}$	Std. Error	t value	Significance
contributes_to_successful_completion	0.6552	0.3610	1.814712E+00	0.0712
making_self_clear	0.9470	0.6080	1.557502E+00	0.1211
engaged_in_game	0.7091	0.3847	1.843187E+00	0.0669
planning_what_to_say	0.3636	0.5756	6.316937E-01	0.5284
gives_encouragement	0.4051	0.3482	1.163506E+00	0.2461
difficult_for_partner_to_speak	0.5287	0.2515	2.101960E+00	0.0369
bored_with_game	-0.0036	0.4106	-8.663987E-03	0.9931
dislikes_partner	0.5307	0.3889	1.364514E+00	0.1741
F0_MEAN	$\widehat{\beta_2}$	Std. Error	t value	Significance
contributes_to_successful_completion	0.9752	0.3058	3.188448E+00	0.0017
making_self_clear	0.6998	0.3907	1.791239E+00	0.0749
engaged_in_game	0.8538	0.2773	3.078945E+00	0.0024
planning_what_to_say	0.6430	0.5363	1.198966E+00	0.2321
gives_encouragement	0.0006	0.3885	1.577445E-03	0.9987
difficult_for_partner_to_speak	-0.5323	0.3835	-1.388190E+00	0.1667
bored_with_game	-0.7663	0.2582	-2.968508E+00	0.0034
dislikes_partner	0.0688	0.3808	1.806265E-01	0.8569
F0_MAX	$\widehat{\beta_2}$	Std. Error	t value	Significance
contributes_to_successful_completion	0.7628	0.4381	1.741129E+00	0.0833
making_self_clear	0.6718	0.4129	1.626984E+00	0.1054
engaged_in_game	0.5308	0.3776	1.405582E+00	0.1615
planning_what_to_say	0.0489	0.4210	1.161167E-01	0.9077
gives_encouragement	0.4724	0.5464	8.647145E-01	0.3883
difficult_for_partner_to_speak	-0.3208	0.2821	-1.136927E+00	0.2570
bored_with_game	-0.2584	0.3764	-6.865032E-01	0.4933
dislikes_partner	0.1249	0.3884	3.216226E-01	0.7481

Fig. 5.2: Tablas con los resultados de la regresión de efectos fijos sobre el va,or absoluto de *entrainment* para ENG_MAX, ENG_MEAN, 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. Las columnas resaltadas corresponden a aquellas significantes, con diferentes matices de gris según $p < 0{,}10$, $p < 0{,}5$, o $p < 0{,}01$

NOISE_TO_HARMONICS_RATIO	$\widehat{\beta_2}$	Std. Error	t value	Significance
contributes_to_successful_completion	0.7271	0.3439	2.114275E+00	0.0358
making_self_clear	1.3576	0.3613	3.758007E+00	0.0002
engaged_in_game	0.1270	0.3431	3.702043E-01	0.7117
planning_what_to_say	-0.1625	0.4264	-3.811856E-01	0.7035
gives_encouragement	0.7665	0.4860	1.577201E+00	0.1165
difficult_for_partner_to_speak	-0.1683	0.3400	-4.951813E-01	0.6211
bored_with_game	0.5527	0.3084	1.792251E+00	0.0747
dislikes_partner	0.3457	0.3279	1.054410E+00	0.2931
PHONEMES_AVG	$\widehat{\beta_2}$	Std. Error	t value	Significance
contributes_to_successful_completion	0.5557	0.3577	1.553747E+00	0.1220
making_self_clear	0.7598	0.5085	1.494093E+00	0.1369
engaged_in_game	0.2440	0.2586	9.438356E-01	0.3465
planning_what_to_say	0.3614	0.5174	6.984626E-01	0.4858
gives_encouragement	0.0604	0.3829	1.576928E-01	0.8749
difficult_for_partner_to_speak	-0.6264	0.3374	-1.856257E+00	0.0650
bored_with_game	-0.0158	0.3204	-4.921947E-02	0.9608
dislikes_partner	0.0975	0.3137	3.108070E-01	0.7563
SYLLABLES_AVG	$\widehat{\beta_2}$	Std. Error	t value	Significance
contributes_to_successful_completion	0.2451	0.3663	6.692398E-01	0.5042
making_self_clear	0.7934	0.4094	1.937743E+00	0.0542
engaged_in_game	0.4956	0.3642	1.360687E+00	0.1753
planning_what_to_say	0.4429	0.5189	8.535430E-01	0.3945
gives_encouragement	0.2363	0.4192	5.637211E-01	0.5736
difficult_for_partner_to_speak	0.1856	0.3481	5.332959E-01	0.5945
bored_with_game	-0.2909	0.3606	-8.067536E-01	0.4208
dislikes_partner	0.1768	0.3452	5.120454E-01	0.6092
SOUND_VOICED_LOCAL_JITTER	$\widehat{\beta_2}$	Std. Error	t value	Significance
contributes_to_successful_completion	0.5770	0.3759	1.534821E+00	0.1265
making_self_clear	0.5057	0.4881	1.036143E+00	0.3015
engaged_in_game	0.4972	0.2515	1.977130E+00	0.0495
planning_what_to_say	-0.0417	0.4628	-9.000210E-02	0.9284
gives_encouragement	-0.0160	0.3502	-4.554031E-02	0.9637
difficult_for_partner_to_speak	-0.2788	0.3126	-8.917922E-01	0.3737
bored_with_game	0.1233	0.3155	3.906725E-01	0.6965
dislikes_partner	-0.1171	0.2788	-4.198582E-01	0.6751
SOUND_VOICED_LOCAL_SHIMMER	$\widehat{\beta_2}$	Std. Error	t value	Significance
contributes_to_successful_completion	0.3745	0.2754	1.359709E+00	0.1756
making_self_clear	-0.0097	0.3821	-2.544762E-02	0.9797
engaged_in_game	0.2434	0.2881	8.449092E-01	0.3993
planning_what_to_say	-0.6040	0.4735	-1.275476E+00	0.2037
gives_encouragement	0.3638	0.2057	1.768094E+00	0.0787
difficult_for_partner_to_speak	0.2707	0.2720	9.952034E-01	0.3209
bored_with_game	-0.3635	0.2772	-1.311203E+00	0.1914
dislikes_partner	-0.1895	0.2667	-7.105564E-01	0.4783

Fig. 5.3: Tablas con los resultados de la regresión de efectos fijos para NOISE_TO_HARMONICS_RATIO, SYLLABLES_AVG, PHONEMES_AVG, SOUND_VOICED_LOCAL_SHIMMER y SOUND_VOICED_LOCAL_JITTER. 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, con diferentes matices de gris según $p < 0{,}10$, $p < 0{,}5$, o $p < 0{,}01$

	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
contributes					
clear	+				+
engaged					
planning					
encourages					
difficult					
bored					
dislikes					

Fig. 5.4: 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 significante y positivo de la pendiente de la regresión de efectos fijos, mientras que − representa significante y negativo

6. CONCLUSIONES Y TRABAJO FUTURO

15.Escribir conclusiones!

7. APÉNDICE

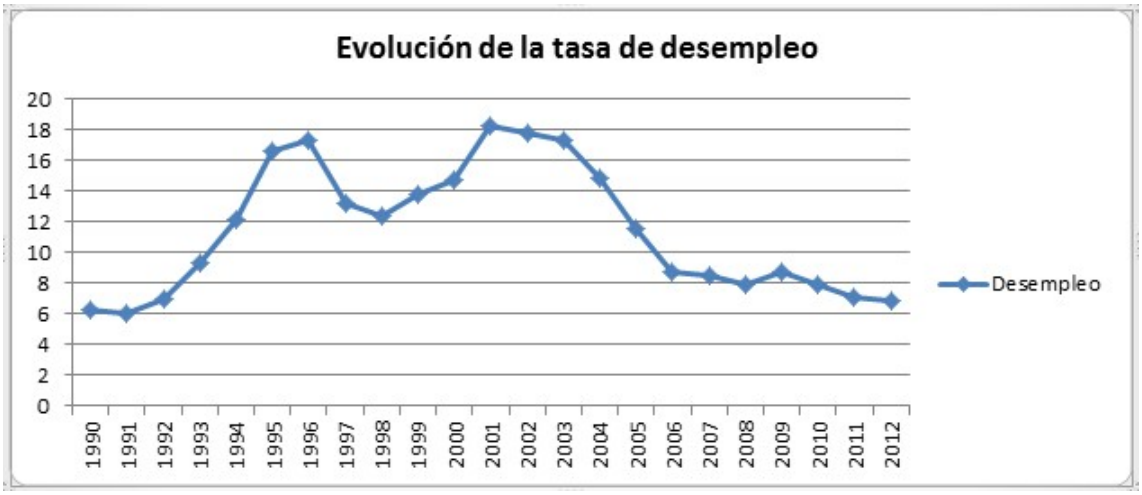

Fig. 7.1: Gráfico de serie de tiempo de la evolución del desempleo en Argentina

7.1. Series de Tiempo

16.Mandar esto a un apéndice!

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, como por ejemplo Economía, Ciencias de la Atmósfera, y otros.

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 7.1 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 esto 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.

7.1.1. Procesos estocásticos

Definición 1. Una 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 $X_1, X_2, \ldots$

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.

7.1.2. 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 $t_1, \ldots, t_n$ y para un desplazamiento $\tau \in \mathbb{N}$ tenemos que

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

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

Se derivan como propiedades que, para todo X_t y cualquier desplazamiento τ

$$E[X_t] = E[X_{t+\tau}] \tag{7.1}$$
$$Var[X_t] = Var[X_{t+\tau}] \tag{7.2}$$
$$Cov(X_s, X_t) = Cov(X_{s+\tau}, X_{t+\tau}) \tag{7.3}$$

Las ecuaciones 7.1 y 7.2 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 7.1, 7.2, 7.3

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

8. BIBLIOGRAFÍA

Bibliografía

[BGT73] Richard Y Bourhis, Howard Giles, and Henri Tajfel. Language as a determinant of welsh identity. *European Journal of Social Psychology*, 3(4):447–460, 1973.

[Bre96] Susan E Brennan. Lexical entrainment in spontaneous dialog. *Proceedings of ISSD*, 96:41–44, 1996.

[BSD95] Judee K Burgoon, Lesa A Stern, and Leesa Dillman. Interpersonal adaptation: Dyadic interaction patterns. 1995.

[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.

[DJ69] James M Dabbs Jr. Similarity of gestures and interpersonal influence. In *Proceedings of the annual convention of the American Psychological Association*. American Psychological Association, 1969.

[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.

[GBLH15] Agustın Gravano, Štefan Benuš, Rivka Levitan, and Julia Hirschberg. Backward mimicry and forward influence in prosodic contour choice in standard american english. In *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[GP99] Damodar N Gujarati and Dawn C Porter. Essentials of econometrics. 1999.

[Gra09] Agustin Gravano. *Turn-taking and affirmative cue words in task-oriented dialogue*. Columbia University, 2009.

[HPH14] Patrick GT Healey, Matthew Purver, and Christine Howes. Divergence in dialogue. *PloS one*, 9(6):e98598, 2014.

[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+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.

[LBGH15] Rivka Levitan, Štefan Benuš, Agustın Gravano, and Julia Hirschberg. Acoustic-prosodic entrainment in slovak, spanish, english and chinese: A cross-linguistic comparison. In *16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, page 325, 2015.

[LH11] Rivka Levitan and Julia Bell Hirschberg. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. 2011.

[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.