



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

Estudio de la Relación entre la Mimetización y la Confianza en un Sistema de Diálogo Hablado

Tesis de Licenciatura en Ciencias de la Computación

Florencia Savoretti

Director: Agustín Gravano

Codirector: Ramiro H. Gálvez

Buenos Aires, 2018

ESTUDIO DE LA RELACIÓN ENTRE LA MIMETIZACIÓN Y LA CONFIANZA EN UN SISTEMA DE DIÁLOGO HABLADO

En las conversaciones entre los seres humanos se da el fenómeno de “entrainment”, también llamado “mimetización” en español, que es la tendencia de una persona a imitar tanto los gestos y posturas como los aspectos acústicos y prosódicos del habla del otro interlocutor. Para este trabajo nos enfocamos en el fenómeno de entrainment acústico-prosódico, adaptando solo la tasa de habla y manteniendo tono, volumen, entre otros. Por otro lado, podemos notar que en una interacción humano-computadora el usuario considerará esa conversación como más “natural” cuanto más se parezca a una conversación interpersonal. Esto nos llevó a plantearnos si el hecho de replicar este fenómeno en una interacción humano-computadora tendría el mismo efecto. Este trabajo se propone entonces responder experimentalmente si podemos afirmar que existe una relación entre el entrainment acústico-prosódico y la confianza generada en el hablante. El experimento se realizó utilizando un juego de cartas que los sujetos debieron jugar a través del habla. El sujeto debía elegir un avatar al cual pedirle un consejo a lo largo de sucesivos turnos. Cada uno de los avatares adoptó una política de adaptación diferente. Luego, de acuerdo a las elecciones del sujeto se estimó el efecto en la confianza provocado por la política de adaptación tomada por cada avatar. Al comparar las políticas de ‘entrainment’ –adaptarse directamente a la tasa de habla del usuario– y de ‘no-effect’ –no adaptar la tasa de habla en absoluto– no se observaron preferencias estadísticamente significativas. En cambio, al contraponer la política de ‘entrainment’ con la de ‘dis-entrainment’ –adaptarse contrariamente a la tasa de habla del usuario–, esta última marcó una leve preferencia por sobre la primera.

Palabras claves: Entrainment, Mimetización, Tasa del habla, Sistema de diálogo hablado, Prosodia.

STUDY OF THE RELATIONSHIP BETWEEN ENTRAINMENT AND TRUST IN A SPOKEN DIALOGUE SYSTEM

In conversations between humans there is a phenomenon known as “entrainment”, which is the tendency of a person to mimic gestures and postures, such as the acoustic and prosodic aspects of speech, of their interlocutor. In this work we focus on acoustic-prosodic entrainment – specifically on the adaptation of speech rate. In human-computer interaction, we notice that the user will consider the conversation as more “natural” the more it resembles an interpersonal conversation. This leads us to consider whether replicating this phenomenon in human-computer interaction would have the same effect. This work is then proposed to respond experimentally whether there is a relationship between acoustic-prosodic entrainment and the degree of trust in the system’s skills developed by the user. The experiment was carried out using a card game that subjects played on a computer through their speech. Subjects had to choose an avatar to ask advice from in successive turns. Each of the avatars adopted a different adaptation policy. Then, according to the subject’s choices, the effect on trust caused by the adaptation policies was estimated. When comparing the ‘entrainment’ policy (adapting directly to the user’s speech rate) and the ‘no-effect’ policy (not adapting the speech rate at all) we observe no statistically significant preferences. On the other hand, by contrasting the policy of ‘entrainment’ with that of ‘dis-entrainment’ (adapting contrary to the user’s speech rate) the latter marked a slight preference over the first.

Keywords: Entrainment, Speech-rate, Spoken Dialog System, Prosody.

AGRADECIMIENTOS

No hubiera conseguido llegar hasta aquí si no fuera por tantas personas que me fueron acompañando durante todos estos años de carrera. Quiero agradecer a todos los que compartieron conmigo días y noches de estudio, otros también trabajo. A ellos: Brian, Hernan, Cele, Dami M., Mati R., Nati, Gabi, Dami B., Maxi, Tom, Caro, Mati Lopez.

A todo el cuerpo docente del departamento, que fueron como una gran familia durante los primeros años de la carrera, donde no solo fueron mis docentes sino también mis pares.

A mi familia de origen, que me dió todo lo necesario para poder llegar hasta aquí.

A mi nueva y hermosa familia. A Gonzalo, mi gran compañero de vida, que me acompañó y apoyó siempre de la mejor manera. A Emma, quien evidentemente con su maravillosa llegada a mi vida movilizó lo necesario para que pudiera presentar este trabajo.

Principalmente quiero agradecer a mis directores Agustín y Ramiro, quienes tuvieron una paciencia sobrehumana para conmigo y mi falta de constancia en este trabajo y me empujaron a finalizarlo.

Índice general

1. Introducción	1
1.1. Motivación	1
2. Método	4
2.1. Descripción del juego	4
2.2. Implementación	5
2.2.1. Interfaz gráfica	6
2.2.2. Funcionamiento del juego	7
2.3. Políticas de adaptación	8
2.4. Diseño del experimento	10
2.4.1. Estructura de una sesión	10
2.4.2. Condiciones de corrida	12
2.4.3. Mediciones del experimento	12
3. Resultados	13
3.1. Elecciones de políticas de adaptación	13
3.2. Exploración vs. explotación	15
3.3. Preferencias reportadas por los sujetos	17
4. Conclusiones	19
4.1. Aportes, limitaciones y trabajo futuro	19
5. Anexo I	21

1. INTRODUCCIÓN

En los **sistemas de diálogo hablado** (SDS - Spoken Dialog Systems) parte del éxito de la experiencia del usuario con los mismos, está supeditado a la **naturalidad** con la que se da la interacción. Esta hipótesis refuerza la idea de que los SDS que replican los comportamientos observados en las conversaciones humano-humano deberían generar una mejor experiencia en el usuario. Siguiendo este razonamiento, uno tendería a pensar que el tono, el volumen y la velocidad en la que se dicen las cosas, i.e. prosodia, siendo una de las componentes esenciales en la comunicación entre humanos, también jugará un rol importante en los diálogos humano-computadora. Por eso, la tendencia es replicar en los SDS los aspectos detectados en las conversaciones entre humanos. [1, 2, 3]

El fenómeno de *entrainment*, también llamado en español *mimetización*, es la propensión de los seres humanos que participan de una conversación a imitar ciertos aspectos del comportamiento del otro interlocutor participante. Esta tendencia se puede observar, tanto en los gestos, si esa conversación es presencial, como en diversos aspectos acústicos y prosódicos del habla ([4, 5, 6, 7] entre otros). Este es uno de los fenómenos que se dan naturalmente en nuestras conversaciones y por eso intentaremos replicarlo en el diálogo humano-computadora.

Para este trabajo, nos enfocaremos en los aspectos acústicos y prosódicos del fenómeno de mimetización. Dentro de las características sociales en las que podría causar impacto, se asocia una correlación entre este fenómeno y la empatía generada en los individuos. Nuestra hipótesis es que también podemos asociar la *confianza* como una consecuencia de la mimetización producida durante la conversación.

En términos generales, nos proponemos responder la siguiente pregunta: **¿existe una relación entre el entrainment acústico-prosódico y la confianza generada en el hablante?**. Para esto, apuntamos a reproducir un experimento que ya ha sido realizado en Inglés [8].

El experimento se realizará con la utilización de un juego en donde se fuerce la interacción humano-computadora a través del habla de manera bidireccional –entre el jugador y dos personajes dentro del juego–, eliminando hasta donde sea posible todas las otras variables que afecten a esta interacción. De esta manera, podremos evaluar la **confianza** del usuario obtenida hacia los dos personajes distintos y sacar conclusiones sobre la consecuencia en el nivel de confianza generada en el sujeto provocada por la política de mimetización de cada uno de los avatares.

Tomando la definición de Mayer [9], la confianza que tenemos en alguien tiene tres componentes: en sus habilidades, en su benevolencia y en su integridad. En nuestro caso, como queremos analizar la confianza hacia un avatar de un SDS, nos enfocamos solamente en la primera de esas componentes.

1.1. Motivación

¿Qué es el fenómeno de mimetización?

En las interacciones humanas, siempre existe algún tipo de coordinación: ésta puede ser de dos tipos: deliberada y determinada por acuerdos sociales –como por ejemplo los

saludos protocolares, o los movimientos en una danza— o bien no intencional y librada de lineamientos —como puede suceder en una conversación casual— sin embargo, más allá de los detalles, la coordinación está presente a pesar de no contar con reglas preestablecidas. En cada tipo de interacción esa coordinación abarca diferentes aspectos, que podría ser con movimientos físicos, gestos o verbalmente acordada.

Una conversación hablada entre dos personas no es la excepción, y también nos encontramos con los aspectos en que esta conversación se coordina. Incluso, si quitamos el foco de los factores sociales que pueden estar influyendo (por supuesto no se dará en los mismos términos la conversación entre un empleado y su jefe que la de un padre con su hijo) la coordinación se produce de todas formas. La coordinación de la que hablamos se puede establecer a través de movimientos, del uso del léxico o bien, aspectos acústicos y prosódicos.

El efecto de *mimetización* tiene muchas aristas, así como los aspectos que mencionamos anteriormente. Una situación frecuente que podría asociarse con este efecto, es cuando hablando con alguien que tiene una tonada regional —un aspecto prosódico— que a mediano plazo se nos va “pegando” la tonada. Esto es una ejemplificación clara de este fenómeno. Algunas otras facetas en las que este fenómeno se puede observar, no son tan evidentes si uno no está atento a ellos; los gestos o posturas que toma el hablante cuando la conversación es presencial, muchas veces son imitados por el otro participante. Por ejemplo, es común que si una de las dos personas está sentada con las piernas cruzadas, el otro participante también tome esa postura.

También nombrado “el efecto camaleón” [10], se refiere a la mímica no consciente de las posturas, gestos, expresiones faciales, y otros comportamientos de uno de los sujetos intervinientes, de tal manera que el comportamiento de uno de forma pasiva e inintencionada cambia para que coincida con la de otros en el propio entorno social. Y sugiere que en una conversación humano-humano, los sujetos tienden a volverse cada vez más parecidos en sus comportamientos a lo largo de la conversación.

¿Por qué nos interesa este fenómeno?

La *mimetización* en la bibliografía sobre interacciones humano-humano nos demuestra tener notoria correlación con variables sociales de manera positiva: la asociamos al aumento de la empatía con el interlocutor, a la mejora en resultados de negociaciones, a la percepción de naturalidad, a mejores reacciones a discursos políticos, etc. [11, 12, 13, 14, 15, 16]

Esto nos sugiere que, si estos resultados fueran también extensibles a las interacciones de tipo humano-computadora, podríamos lograr, con un agente que dinámicamente se ajuste al comportamiento del humano, que el humano perciba la conversación como más natural y posiblemente logremos que este reaccione más favorablemente en esa interacción.

¿Por qué reiteramos el experimento en Español?

El método utilizado en este experimento, básicamente es una réplica del utilizado por Rivka Levitan [8] en el idioma Inglés. Implementamos un sistema de diálogo hablado que modifique características de su habla en función a las características del hablante.

Pretendemos estudiar experimentalmente la causalidad entre el fenómeno de mimetización y la confianza generada en el sujeto. Como el sentimiento de confianza entra en juego en esta pregunta, los aspectos socioculturales juegan un papel importante en el experimento. Y estos aspectos también van a afectar a la forma del habla, por ejemplo, hay

regiones en que se suele hablar más rápido que en otras.

Así que no sólo es el idioma lo que estamos cambiando, sino también hay que tener en cuenta estos aspectos inherentes en los sujetos participantes. Es decir, que los resultados obtenidos en un idioma/cultura no son inmediatamente expansibles a otro, es por eso que nos interesaba reiterar el experimento en Español.

El protocolo empleado en los experimentos de esta tesis fue evaluado y aprobado por el Comité de Ética del Centro de Educación Médica e Investigaciones Clínicas (CEMIC) “Norberto Quirno”, el 18 de julio de 2014, con validez hasta el 31 de agosto de 2017.

A continuación, describiremos el diseño del experimento, anteponiendo la presentación del juego de cartas utilizado y las políticas de adaptación con los que se ha llevado a cabo el experimento. Luego, veremos los resultados de la ejecución del experimento y por último, las conclusiones del trabajo realizado.

2. MÉTODO

2.1. Descripción del juego

El juego utilizado para llevar a cabo el experimento es una modificación del clásico juego de cartas *Go Fish*, al cual dimos por llamar “A freir churros” intentando otorgarle una impronta local. A diferencia del juego original, en el que intervienen dos o más jugadores, en esta versión, el sujeto jugará directamente con la computadora actuando como su oponente. Sin embargo, la computadora no acumulará puntos, ya que el foco del juego está en los puntos acumulados por el sujeto en cada partida.

La característica más importante del juego se basa en que el jugador sólo podrá actuar a través de lo que le indiquen las consejeras –también llamadas avatares– de tal manera que, más que un consejo, es una orden. Como adelantábamos, tampoco competirá en realidad contra su oponente, sino que para el sujeto, el objetivo del juego será sumar la mayor cantidad posible de puntos, a través de formar todos los *sets* (conjunto de 4 cartas del mismo número) que le sea posible.

Inicialmente, se le reparten siete cartas a cada jugador (el sujeto y la computadora). Una vez hecho esto, comienza el primer turno: el sujeto puede elegir a alguna de las dos consejeras y solicitarle su consejo, el cual consiste en sugerirle una carta para pedirle al contrincante (por ejemplo, puede sugerirle que pida un 3). El sujeto deberá obligatoriamente seguir ese consejo ya que, como veremos más adelante, la implementación del juego intencionalmente no le permite omitir el consejo. Una vez realizado el pedido de cartas, su contrincante (la computadora) está obligado a darle todas las cartas que tenga de ese número. Aquí pueden darse tres situaciones:

1. La computadora le da sus cartas y el sujeto pasa a tener los cuatro palos de ese número de carta, es decir, que ha formado un *set*. En nuestro ejemplo, el sujeto forma un set con $3\heartsuit$, $3\diamondsuit$, $3\clubsuit$ y $3\spadesuit$

En este caso, el sujeto gana *10 puntos por cada carta* que reciba de su contrincante y *100 puntos más* por haber formado un *set*.

2. La computadora le da sus cartas, pero aún así el sujeto no consigue completar el *set* de ese número.

En este caso, el sujeto gana *10 puntos por cada carta* que recibe y conserva las cartas.

3. La computadora no tiene ninguna carta del número que el sujeto pidió. En este caso, se lo penaliza y la computadora lo manda a “Freir Churros”.

Este caso lo obliga a levantar una carta del mazo y *perder 50 puntos*.

Seguido a este paso, la computadora devuelve todas sus cartas restantes al mazo, las mismas se van a mezclar y se le volverá a repartir a la computadora siete nuevas cartas, con esto finaliza el turno.

En total, la partida dura 15 turnos y una vez cumplidos, se acaba, mostrándole al sujeto el puntaje obtenido en esa partida.

Es importante destacar que **en cada turno, el jugador debe optar por una de los avatares** a la que le solicitará su consejo.

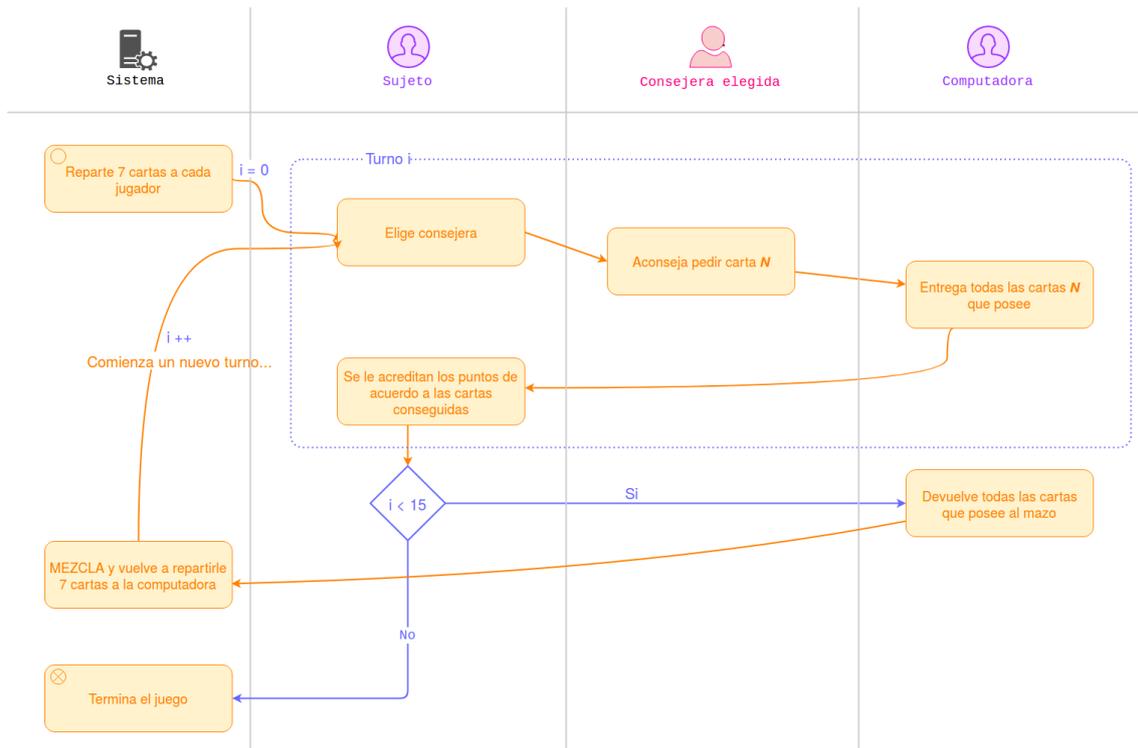


Fig. 2.1: Diagrama de flujo del juego "A freir churros"

Para más claridad, mostramos el flujo del juego de manera gráfica, en la Figura 2.1:

2.2. Implementación

Para la implementación del juego ¹, nuestra primer opción fue elegir un lenguaje multiplataforma que nos permitiera no solo correrlo en diferentes entornos, sino también, reutilizar el código en los diversos entornos que presentan tanto los investigadores del área, como los colaboradores de otras universidades. De esta manera, conseguimos exitosamente que las líneas de investigación que siguieron pudieran continuar el trabajo sin esta barrera.

Vale mencionar, que desarrollar el juego desde cero no fue el primer enfoque, sino que antes intentamos recuperar el código fuente del juego original desarrollado por Rivka Levitan para su tesis [8]. Sin embargo, no lo logramos debido a que estaba desarrollado en Objective-C para ejecutar sobre MacOS y eso no nos permitió ejecutarlo directamente en nuestros entornos, ni logramos adaptarlo.

Por otro lado, varios investigadores del área que incursionaron en el desarrollo de juegos interactivos resolvieron el manejo multimedia con la utilización de un framework llamado *pygame*. Con lo cual, contemplando estos motivos, la elección de *Python* como lenguaje de programación parecía acertada y verdaderamente nos sirvió para cumplir nuestro propósito.

¹ El código del juego queda disponible para compartir (contactar a los directores de la tesis).

2.2.1. Interfaz gráfica

Como hemos mencionado con anterioridad, una de las principales características del experimento es que el juego requiere la interacción con las consejeras. En la Figura 2.2 presentamos a los avatares utilizados para cumplir ese rol. Es importante destacar, que a fin de evitar preferencias de afinidad con la apariencia o el nombre del avatar, las políticas de adaptación se fueron aplicando a una y otra consejera indistintamente, para evitar que los resultados de preferencia sean provocados por otra causa que no fuera la política de adaptación.



Fig. 2.2: Avatares: Eugenia a la izquierda, Amanda a la derecha

En cuanto a la interfaz gráfica del juego, que tuvieron que utilizar los sujetos durante el experimento, se pueden ver en la Figura 2.3 algunas capturas de pantalla, con cada uno de los avatares mientras dan su consejo. En estas imágenes también podemos apreciar que el sujeto puede ver constantemente en pantalla las frases que el juego acepta, así como también los nombres de las consejeras junto con sus avatares para tener como recordatorio para que pueda concentrarse en el juego y no en tener que recordar las reglas con las que lo estamos haciendo jugar.



Fig. 2.3: Imágenes del juego "A freir churros" con ambas consejeras

2.2.2. Funcionamiento del juego

El juego es interactivo, pensado para que, durante una partida, el sujeto deba interactuar con este mediante el habla obligatoriamente. Con ese fin, el juego sólo cuenta con un botón que se debe mantener presionado mientras se le habla a la consejera. Este tipo de botones son denominados PTT – push to talk, como el que se utiliza en WhatsApp para enviar audios en vivo, lo que nos pareció que sería una interfaz conocida por la mayoría de los sujetos. Este botón es el que le permite al sujeto transmitir su intención verbalmente.

A lo largo de un turno, lo primero que sucede es que el sujeto debe pedirle un consejo a uno de los dos avatares. Para ello, no tiene otra forma de hacerlo que no sea nombrándolas y haciéndoles una de las preguntas estipuladas para la conversación. Las posibles preguntas son cinco, lo que nos deja 20 posibles frases a reconocer, compuestas por: 5 preguntas, 2 nombres de consejeras y 2 posiciones para decir el nombre, por ejemplo: “Amanda, ¿qué hago ahora” ó “¿Qué carta pido, Eugenia?”. Para mostrar con mayor claridad estas 20 posibilidades, mostramos la definición de la gramática utilizada:

$$\begin{aligned} \$ &\longrightarrow N, Q? \mid Q, N? \\ N &\longrightarrow \text{'Eugenia'} \mid \text{'Amanda'} \\ Q &\longrightarrow \text{'Qué carta pido'} \mid \text{'Qué te parece que haga'} \mid \text{'Qué pensás que conviene'} \mid \\ &\text{'Qué me recomendás'} \mid \text{'Qué hago ahora'} \end{aligned}$$

Una vez recibida la consulta del sujeto, se utiliza *PocketSphinx* [17] sobre el audio capturado como reconocedor de habla – más comúnmente conocido como ASR (del inglés, *automatic speech recognition*)– para el cual se ha definido la gramática mencionada anteriormente. El ASR luego de procesar la frase, si resulta ser una cadena válida dentro de la gramática nos devuelve como resultado la frase reconocida y el tiempo que ha demorado en decirla, es decir, la duración del habla del sujeto.

La tasa del habla, ó *speech-rate*, es la velocidad del habla en un discurso oral. Puede medirse en unidad de palabras por minuto, o para ser más preciso, en sílabas por segundo. Es decir, es la cantidad de elementos fónicos en un período de tiempo determinado.

Con el resultado otorgado por el ASR, podemos calcular directamente la tasa del habla. De hecho, para esto se utiliza una fórmula muy sencilla:

$$\text{tasa del habla} = \frac{\text{cant de sílabas}}{\text{duración del habla}}$$

Por ejemplo, si el sujeto dice la frase “Amanda, Qué carta pido?” y demora 1360 ms en decirla podemos calcular:

Duración del habla = 1360 ms = 1,36 segundos

Cantidad de sílabas = 3 (Amanda) + 1 (qué) + 2 (carta) + 2 (pido) = 8 sílabas

Luego,

$$\text{tasa del habla} = \frac{8 \text{ sílabas}}{1,36 \text{ seg}} = 5,8823 \text{ sílabas por segundo.}$$

Una vez que el consejo fue solicitado, y la frase fue comprendida de manera exitosa por el reconocedor, el juego hace que la consejera elegida –reconocida en la frase– le hable

al sujeto para decirle la carta que debe solicitar al oponente (la computadora). Para hacer que la consejera hable, se utiliza la tasa de habla calculada según la estrategia de habla que ha sido establecida para ese avatar.

Para darles voz a las consejeras del experimento, utilizamos un sistema de síntesis de habla (o TTS, del inglés *text to speech*) en español argentino, desarrollado como parte de la Tesis de Licenciatura de Luisina Violante [18]. Dicho sistema está basado en el framework de código abierto MaryTTS² y fue construido a partir del corpus SECyT de alocuciones cortas leídas por una locutora profesional [19].

El *pitch* o *tono* de la voz es una variable acústica basada en la frecuencia fundamental de la señal, y se lo considera uno de los principales atributos perceptuales del habla. Por eso, y debido a que contamos con un único sistema de síntesis de habla, decidimos generar las voces de las consejeras con dos niveles de pitch diferentes para distinguirlas, consideramos que esto le otorgó personalidad a las voces de cada una de las consejeras.

2.3. Políticas de adaptación

Dado un segmento del habla, es posible medir su velocidad, y si tomamos varias muestras de segmentos de un mismo sujeto, podremos determinar cuál es su velocidad promedio de habla. Esto nos da una clara referencia de qué discurso es rápido para ese sujeto y cuál es lento.

Algunos términos que se utilizarán más adelante a tener en cuenta:

<code>rateSist</code>	La velocidad que utiliza por defecto el sistema
<code>rateUsrBase</code>	La velocidad promedio de habla del sujeto medida en el juego anterior
<code>rateUsrAct</code>	La velocidad del habla calculada para la consejera actualizada luego de aplicada la estrategia
<code>rateNew</code>	La medida de la tasa del habla del sujeto en la frase grabada durante el turno actual

Para el experimento, hemos determinado 3 posibles estrategias, o *políticas de adaptación*, que podrían utilizar las consejeras como métodos para adaptar su voz a la del sujeto participante. A continuación, detallamos cada una de ellas:

Entrainment Esta estrategia consiste en actualizar la velocidad del habla de la consejera de acuerdo a la variación de la tasa de habla del sujeto. Sencillamente, si el sujeto habla más rápido que antes, la consejera también le responderá más rápido. Esto se puede representar con la siguiente ecuación:

$$rateUsrAct = \frac{rateNew - rateUsrBase}{rateUsrBase}$$

Dis-entrainment Esta estrategia consiste en actualizar la velocidad del habla de la consejera de manera contraria al cambio de tasa del habla del sujeto. Es decir, que si el sujeto habla más rápido que antes, la consejera lentifica su respuesta. Esto se puede representar con la siguiente ecuación:

$$rateUsrAct = \frac{rateUsrBase - rateNew}{rateUsrBase}$$

² <http://mary.dfki.de>

No-effect Esta estrategia consiste en mantener siempre la velocidad promedio del sistema para la generación de la voz de las consejeras. La consejera siempre mantendrá la misma tasa de habla. Esto se traduce al uso de esta ecuación:

$$rateUsrAct = rateSist$$

En la próxima sección veremos cómo fueron utilizadas estas estrategias.

Antes de continuar, una de las preguntas que nos debemos hacer es si pudimos exponer a los sujetos a lo que pretendíamos. Es decir, ¿la tasa del habla del avatar, en verdad se adaptó en la medida que el sujeto variaba su discurso?.

En palabras más técnicas, lo que queremos confirmar es cuando al TTS le pedimos que sintetice con una variación de $+/- X\%$, efectivamente lo hace en ese porcentaje. Para responder esta pregunta, sintetizamos las siguientes 9 frases:

1. “yo te sugiero pedir un siete”
2. “yo te sugiero pedir una cu de queso”
3. “yo te sugiero pedir un nueve”
4. “te aconsejo buscar un as”
5. “te aconsejo buscar un dos”
6. “te aconsejo buscar un cuatro”
7. “te recomiendo pedir una jota”
8. “te recomiendo pedir una ca de kilo”
9. “te recomiendo pedir un diez”

Estas frases se sintetizaron utilizando 31 valores diferentes de modificación de speech-rate, a saber, esos valores son los porcentajes desde -15% hasta $+15\%$ de uno en uno (incluyendo 0%).

Luego, a cada uno de esos 279 wavs obtenidos ($9 \text{ frases} \times 31 \text{ porcentajes}$) se les midió la tasa de habla (speech-rate) y comparamos el valor de speech-rate que se obtendría con una modificación de $+0\%$. Con estos valores, graficamos boxplots para los valores obtenidos (cada boxplot tiene 9 observaciones, una por oración sintetizada) y es lo que exponemos en la Figura 2.4.

Observemos que idealmente si las frases sintetizadas tuvieran exactamente la misma tasa de habla que le solicitamos obtendríamos una diagonal con inclinación de 45° y boxplots angostos como sucede cerca del 0% . Pero no es el caso, efectivamente a medida que variamos el speech-rate más lejos del 0, obtenemos variaciones en las direcciones deseadas, pero menores a la magnitud deseada. Esto es una característica del TTS y se ha convertido en una limitación de este trabajo. Como trabajo a futuro, sería pertinente evaluar otros sistemas de TTS y correr el experimento con alguno que efectivamente sintetice en la magnitud que deseamos.

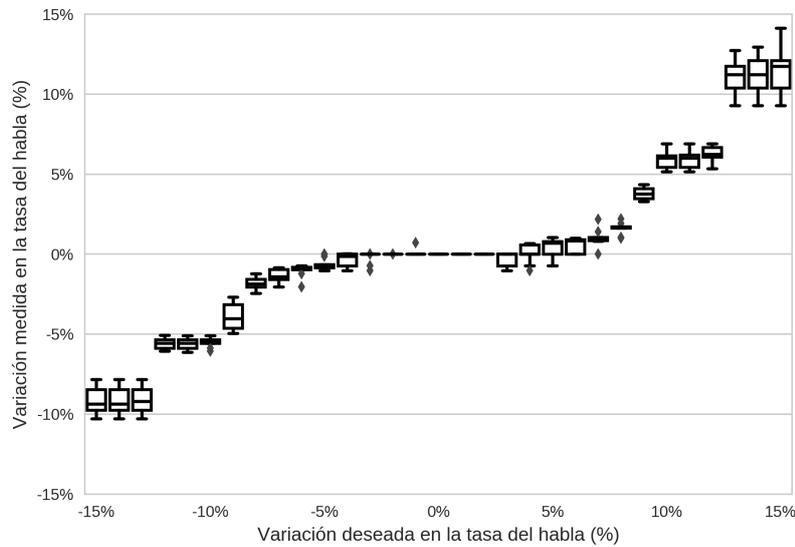


Fig. 2.4: Variaciones en la tasa del habla

2.4. Diseño del experimento

El experimento se basa en la interacción de un sujeto con un juego de cartas llamado “A freir churros”, la modificación del clásico *Go Fish* explicada en la sección 2.1.

En una sesión del experimento (la exposición de un sujeto frente al juego durante 3 partidas), cada consejera mantiene la política de adaptación durante toda la sesión.

El sujeto debe interactuar mediante el habla con las consejeras, y ellas le responderán con su tono de voz característico, pero adaptando o no la tasa de habla de acuerdo a la política de adaptación utilizada por cada una de ellas en esa corrida. De esta manera, podremos observar si una política de adaptación es más elegida por los sujetos.

2.4.1. Estructura de una sesión

Como ya explicamos con anterioridad, intentamos determinar si la *confianza* generada en el hablante es afectada de alguna forma por el fenómeno de mimetización. Para ello necesitábamos despejar del experimento todas las posibles variables que pudieran interferir en la interacción hablada de manera que no afectaran al resultado de manera indeseada. Para lograr este objetivo, mostramos cómo llevamos a cabo cada sesión del experimento en las cuales se realizaron siempre los mismos pasos. A continuación pasamos a detallarlos:

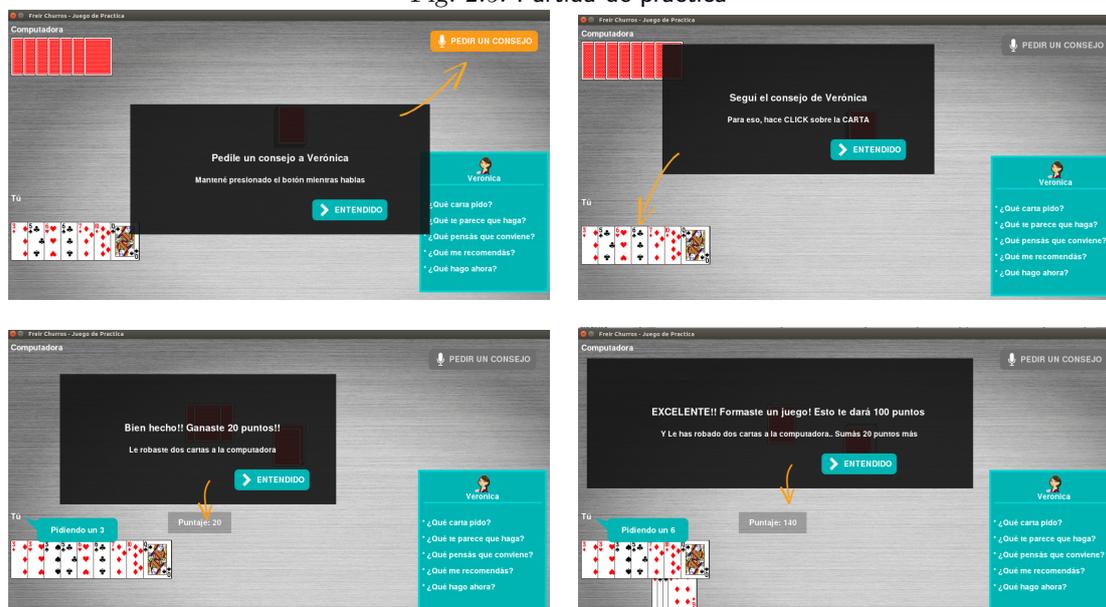
En cuanto llegan los sujetos que participarían de la sesión se los ubicó en las computadoras que utilizarían para el experimento, con todo el equipamiento necesario previamente preparado. Lo primero que hacemos es presentarnos y explicarles verbalmente de qué se tratará el experimento con los aspectos más importantes del juego, haciendo hincapié en la observación de que **cada consejera utiliza su propio algoritmo para dar consejos** y ellos deberían ir descubriendo cuál es el mejor para ganar el juego. Con esto, logramos enfatizar que el sujeto debería esforzarse por determinar en quién confiar más, suponiendo y remarcando que alguna le daría mejores consejos que la otra. Luego se les entregó una

hoja con las instrucciones del juego –véase en sección 5– y se les otorgó suficiente tiempo para que pudieran leerla e hicieran consultas al respecto.

La primer partida de cada experimento se realizó con una sola consejera –distinta de las dos utilizadas para el estudio, tanto el rostro como la voz, que no cambió en ningún momento ni su tono de voz ni su tasa de habla– y sólo 5 turnos de juego, lo cual dimos por llamar *partida de práctica*. Esta partida nos sirvió para explicar los movimientos del juego y los sectores de la pantalla. Se les mostró una serie de pasos estilo tutorial de juego para que ubiquen todas las partes de la pantalla.

La partida de práctica sirvió también para que el sujeto pruebe la comunicación con la consejera y sobre todo, para tomar los primeros 5 audios del sujeto. De esos cinco audios descartamos el primero, pensando en que podría no ser tan natural, y a los otros cuatro los usamos como muestra para determinar su *rateUsrBase* que utilizaríamos luego en las partidas reales. En las imágenes de la Figura 2.5 se pueden visualizar algunas capturas de pantalla de la partida de práctica en donde se aprecian los aspectos recién mencionados.

Fig. 2.5: Partida de práctica



Al terminar la partida de práctica, se les dio un espacio por si surgía alguna duda, y luego se los dejó jugar las otras tres partidas reales. En cuanto terminaron de jugar, se les dió un cuestionario para que lo completen –ver sección 5–

Por último, los sujetos participantes recibieron un premio monetario de acuerdo al desempeño en el juego, cuantificado en relación al puntaje obtenido a lo largo de las tres partidas, el cual les fue promocionado antes de comenzar a jugar. Con este incentivo, intentamos fomentar la intención de ganar el juego, y prestar atención en cada elección, ya que la única manera de jugarlo era a través de la elección de las consejeras y de esa manera, la elección correcta les haría ganar un mayor premio.

2.4.2. Condiciones de corrida

Una sesión del experimento está compuesto por tres partidas del juego que el sujeto debe jugar –de 15 turnos cada uno– posteriores a la partida de práctica.

Al iniciar una sesión del experimento, se configura la condición en que correrá el juego. Se decide entre dos posibles condiciones:

Entrainment vs. Dis-entrainment En esta condición, una consejera utiliza la estrategia de **Entrainment** y la otra, la estrategia **Dis-entrainment**

Entrainment vs. No-effect En esta condición, una consejera utiliza la estrategia de **Entrainment** y la otra, la estrategia **No-effect**

La mitad de las corridas fueron en la condición **Entrainment vs. No-effect** y la otra mitad en la condición **Entrainment vs. Dis-entrainment**. A su vez, se varió además el avatar que realizaba la estrategia de **entrainment**. Es decir, en cada una de las condiciones, la mitad de las veces se adapta **Amanda** y la otra mitad se adapta **Eugenia**. Esto hará que obtengamos resultados para cada una de las condiciones equilibradamente, y lo que evitamos con esto, fue un efecto indeseado de preferencias de avatar.

2.4.3. Mediciones del experimento

En la sesión se medirá cuántas veces el sujeto ha elegido a cada consejera teniendo en cuenta en cada caso la condición en que corre el experimento. Además, inmediatamente al terminar de jugar, el sujeto debe completar un cuestionario de preferencias –ver sección 5– que intentarán reflejar tanto la preferencia de consejera como los sentimientos provocados por las voces de cada una de ellas.

3. RESULTADOS

El experimento fue realizado por 26 sujetos, muestra que a priori consideramos sería suficiente para notar una tendencia, en caso de existir. Los 26 participantes son todos de la zona rioplatense, salvo uno de ellos que es panameño, con una participación de 20 hombres y 6 mujeres. En cuanto a la edad, el promedio es de 28.3 años con un desvío standard de 8.8 años.

De los 26 sujetos que participaron, 14 lo hicieron en la condición 1 (Entrainment vs Dis-entrainment), y 12 sujetos en la condición 2 (Entrainment vs No-effect). Recordemos que en cada sesión, el sujeto juega 3 partidas del juego, las cuales tienen 15 turnos cada uno; eso nos deja 45 interacciones de habla de cada sujeto con el juego. Por lo tanto, serán también 45 oportunidades para decidir entre un avatar y otro.

En base a esas decisiones, estudiamos el efecto que tiene la política de adaptación del avatar. Es decir, cada avatar ha adoptado una política de adaptación –i.e. Entrainment / Dis-entrainment / No-effect–, y con las 45 elecciones de avatar que toma el sujeto veremos cómo estas condiciones afectaron su confianza.

3.1. Elecciones de políticas de adaptación

En los gráficos de la Figura 3.1 podemos observar por cada sujeto la cantidad de elecciones entre una política y la otra, para cada una de las condiciones del estudio. Se debe tener en cuenta, que lo que importa no es la elección entre un avatar y otro, sino que, como los avatares fueron cambiando su política de sujeto en sujeto, se analiza la elección entre distintas políticas de adaptación. Para apreciarlo mejor visualmente, los resultados fueron ordenados en este gráfico de acuerdo a la cantidad de elecciones del sujeto, dejando arriba a los sujetos que más veces eligieron la política de Entrainment. Por ejemplo, para la figura de la izquierda (condición 1) el sujeto representado en la barra superior ha elegido 3 veces al avatar con política de entrainment, y las 42 veces restantes al avatar que aplica la política de dis-entrainment.

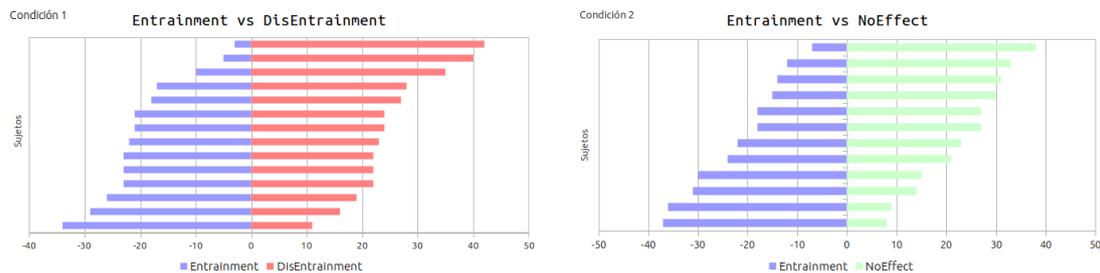


Fig. 3.1: Elecciones de política de adaptación de cada sujeto

En la condición 1 –Entrainment vs Dis-entrainment– se puede observar una leve preferencia por la política de Dis-entrainment. En cambio, en la condición 2 –Entrainment vs No-effect– se puede percibir una diagonal más clara, evidenciando que la cantidad de elecciones es muy pareja entre las dos políticas de adaptación

A su vez, graficamos las elecciones a nivel general, es decir, aglomerando las elecciones de todos los sujetos. Se resume la cantidad de elecciones entre una política y otra en los gráficos de la Figura 3.2, y aquí se aprecia con más claridad lo que mencionamos en gráficos anteriores. Vemos que el gráfico de torta de la condición 2 refleja que las elecciones fueron prácticamente mitad y mitad entre las dos políticas, mientras que en la condición 1 existe una leve preferencia por la política de Dis-entrainment por sobre la de Entrainment

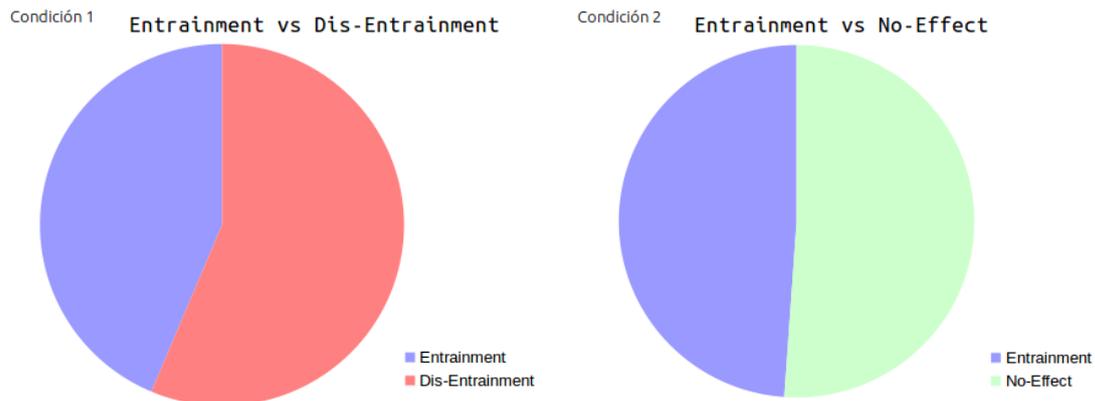


Fig. 3.2: Elecciones de política de adaptación a nivel general

Para medir la significancia estadística de nuestros resultados, hacemos uso de tests binomiales exactos bidireccionales. Estos tests miden qué tan probable es que dado “p” (una probabilidad de éxito dada; es decir, la probabilidad que un evento dado ocurra), se den “n” casos de éxito en un total “N” pruebas. A modo de ejemplo, si se quiere testear que la probabilidad de que una bombilla cualquiera de luz se quemara es de 0,01 (p) y observamos que de una población de 50 bombillas (N) 30 bombillas están quemadas (n), el test indicaría que, dados nuestros datos, es muy improbable que la probabilidad subyacente de que una bombilla se quemara es de 0,01 (arrojaría un bajo p-valor). Para el caso de “A Freir Churros”, queremos descartar que la elección de consejeras fue totalmente al azar. De este modo en nuestros test evaluaremos qué tan consecuente es haber obtenido los resultados obtenidos con que la probabilidad de elegir cada consejera sea igual a 0,5.

Para el caso de la condición 1 observamos que de los 630 elecciones evaluadas, en 275 se eligió a la consejera que seguía una política de entrainment (43,7%). Un test binomial exacto bidireccional indica que la probabilidad de que esto ocurra dada una probabilidad de elegir a la consejera que hace entrainment de 0,5 es menor a 0,001 ($p < 0,001$). De modo que existe evidencia que sugiere que la selección de consejeras no fue totalmente aleatoria.

Para el caso de la condición 2 observamos que de los 540 elecciones evaluadas, en 264 se eligió a la consejera que seguía una política de entrainment (48,9%). Un test binomial exacto bidireccional indica que la probabilidad de que esto ocurra dada una probabilidad de elegir a la consejera que hace entrainment de 0,5 es igual a 0,636 ($p = 0,636$). De este modo, dados nuestros datos, no podemos descartar que las consejeras fueron elegidas de una manera totalmente aleatoria.

3.2. Exploración vs. explotación

El primer análisis realizado sobre los resultados es el inmediato, de manera muy básica, sólo teniendo en cuenta la cantidad de elecciones. Sin embargo, por cómo es el desafío que se le plantea al sujeto, que consiste en decidir cuál es la consejera que lo ayudará a ganar, esperamos que el comportamiento durante el juego tenga dos etapas. Primero el sujeto explora, evaluando las posibilidades y a medida que avanza en el juego, se va inclinando por una de las dos consejeras. Esto es un ejemplo del dilema de exploración vs. explotación en problemas de aprendizaje por prueba y error. Este dilema se da cuando tienes que decidir si sigues tomando la misma decisión que hasta el momento que te hizo llegar a donde estas o cambias y decides explorar nuevas decisiones.

Lo primero que debemos hacer es evaluar si el comportamiento sospechado se da efectivamente. Para esto creamos una variable que toma el valor de 1 si se elige al avatar de Amanda o 0 en caso contrario. Luego, para cada dupla sujeto-juego calculamos la varianza de esta nueva variable. Tengamos en cuenta que una **varianza cercana a 0** significa que siempre se eligió al mismo avatar y una **varianza cercana a 0,25** significa que las elecciones fueron mitad y mitad para cada avatar.

Por último, tomamos el promedio de la varianza calculada para cada juego y esto nos va a decir cuánto han explorado los sujetos en cada uno de los juegos. Esto representa un indicador de la diversidad de las elecciones.

Juego	Varianza promedio
1° juego	0,2123
2° juego	0,1418
3° juego	0,1063

Tab. 3.1: Varianza promedio de elección de avatar por juego

Podemos ver en el Cuadro 3.1 que efectivamente, en el primer juego se exploran más las opciones, y a medida que avanza el experimento, los sujetos se van decidiendo por uno de los dos avatares, ya que el promedio de la varianza baja en cada juego respecto del anterior. Esto nos motiva a repetir el mismo análisis anterior, pero ahora ponderando las decisiones con más peso a medida que se acerca el final del juego.

La manera que utilizamos para calcular esto, fue dándole peso a las decisiones de los sujetos de la siguiente manera, utilizando los pesos 1, 2 y 3 para los juegos según el orden, para dar mayor importancia a las decisiones de la etapa de explotación según lo aprendido en la etapa de exploración:

$$\text{Elecciones}_a = X_1 \times 1 + X_2 \times 2 + X_3 \times 3$$

Siendo:

- **Elecciones_a** la cantidad de elecciones ponderadas de un sujeto por el avatar a;
- **X₁** la cantidad de elecciones por el avatar utilizando política a en el primer juego;
- **X₂** la cantidad de elecciones por el avatar utilizando política a en el segundo juego;
- **X₃** la cantidad de elecciones por el avatar utilizando política a en el tercer juego

De esta forma, en la Figura 3.3 tomemos como ejemplo al primer sujeto –i.e. barra inferior– que eligió 9 veces al avatar con Entrainment y 6 veces al avatar con Dis-entrainment en el primer juego; 10 y 5 veces en el segundo juego; y 15 y 0 en el tercero. Luego, aplicándolo a la fórmula antes mencionada, obtenemos:

$$\begin{aligned} \text{Elecciones}_e &= 9 \times 1 + 10 \times 2 + 15 \times 3 \\ &= 9 + 20 + 45 \\ &= 74 \end{aligned}$$

$$\begin{aligned} \text{Elecciones}_d &= 6 \times 1 + 5 \times 2 + 0 \times 3 \\ &= 6 + 10 + 0 \\ &= 16 \end{aligned}$$

donde,

- *e* es el avatar que aplica la política de Entrainment y
- *d* es el avatar que aplica la política de Dis-entrainment

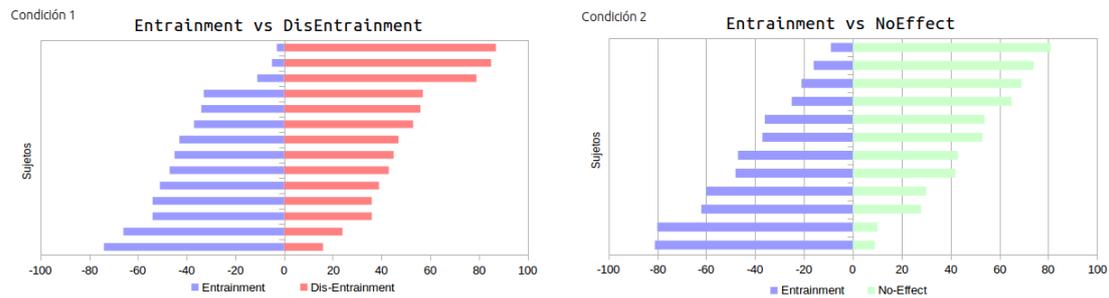


Fig. 3.3: Elecciones de política de adaptación por sujeto **ponderadas**

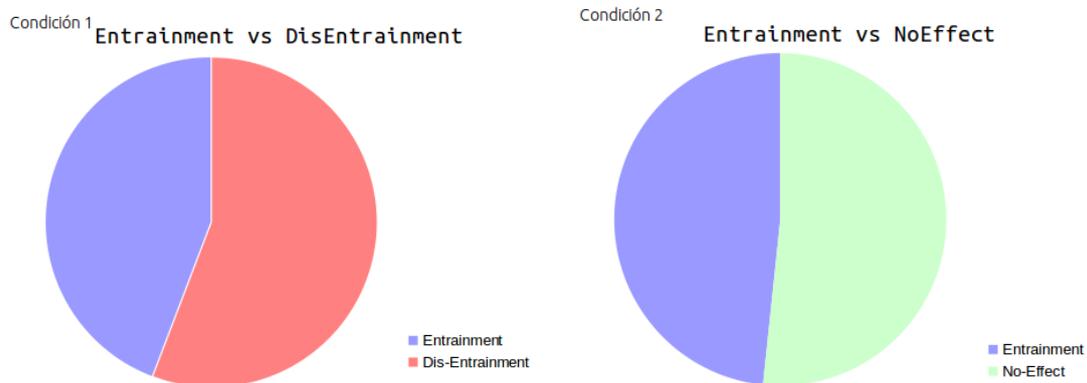


Fig. 3.4: Elecciones de política de adaptación a nivel general **ponderadas**

Efectivamente, en esta nueva manera de ver los datos, los resultados revelan el mismo comportamiento que mencionamos en el primer análisis de los datos –ver Figura 3.3 y 3.4–. Esto nos dice que, por lo general, los sujetos a medida que avanzaban los turnos se iban decidiendo por un avatar. En la condición 1, la tendencia fue inclinarse por la política de Dis-entrainment, sin embargo, en la condición 2, no hubo preferencia en lo absoluto, dejando las políticas de Entrainment y No-effect totalmente emparejadas.

3.3. Preferencias reportadas por los sujetos

Por otro lado, como explicamos anteriormente, les hemos realizado 3 preguntas a los sujetos sobre las voces de los avatares –ver cuestionario en sección 5–, estas son: ¿Cuál consejera te gustó más?, ¿Cuál de las voces te gustó más? y ¿Cuál consejera te dio mejores consejos?

En la Figura 3.5 hemos graficado los resultados. Si bien las respuestas de los sujetos a las preguntas sobre las voces no muestran exactamente el mismo resultado que las elecciones, sí se puede ver que reflejan la misma tendencia. Esto es, en la condición 2 la preferencia de las voces que tomaron las distintas políticas de adaptación se mantuvo dentro de todo pareja; mientras que en la condición 1 se ve una preferencia por la política de Dis-entrainment por sobre la de Entrainment

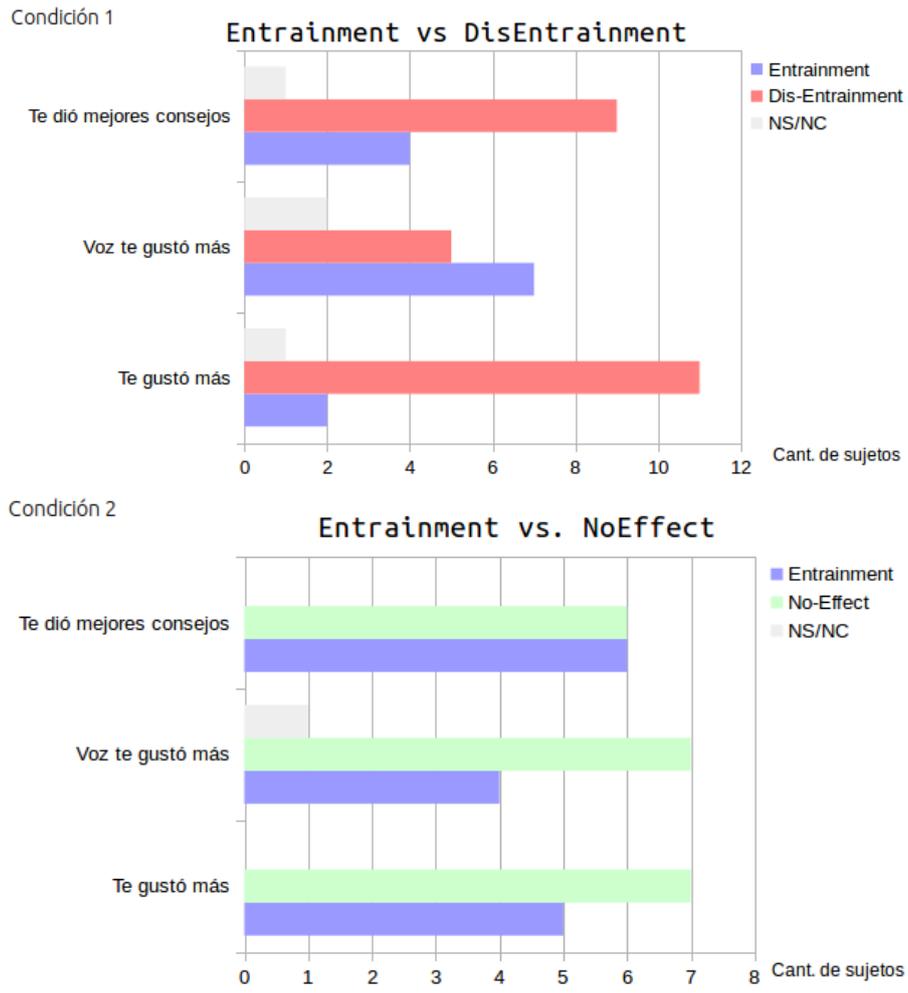


Fig. 3.5: Respuestas a las preguntas sobre las voces

4. CONCLUSIONES

En este trabajo, nos hemos propuesto responder la pregunta: *¿Existe una relación entre el entrainment acústico-prosódico de un sistema de diálogo hablado y la confianza generada en el usuario?*. Para ello diseñamos e implementamos un sistema de diálogo hablado con la capacidad de adaptar dinámicamente su tasa de habla a la del usuario.

El experimento se realizó enfrentando al sujeto a un juego de cartas en el que debía interactuar con dos consejeras implementadas mediante sendos sistemas de diálogo. Cada una de ellas, adoptaría una política de adaptación (de la tasa de habla) con respecto a la tasa de habla del sujeto.

La consigna que se les planteó a los sujetos en todo momento hacía hincapié en que debían averiguar cuál consejera le estaba otorgando mejores recomendaciones; además, la única variable que dejamos librada en la elección que hace cada sujeto por uno de los avatares (a quién decidía pedirle su consejo) era la política de adaptación de su voz. Es importante resaltar que no se han provocado interferencias en el experimento por preferencias de un avatar sobre otro. Esto lo logramos balanceando la cantidad de veces que cada avatar adoptaba una política u otra. Por estos motivos conjuntos, es que las elecciones por un avatar sobre el otro reflejan el efecto que tiene la política de entrainment del avatar sobre el nivel de confianza del usuario.

Las políticas de adaptación que utilizaron las consejeras durante el experimento fueron tres: **Entrainment**, en donde el avatar variaba su tasa de habla proporcionalmente a como lo hacía el sujeto; **Dis-entrainment**, en donde el avatar variaba su tasa de habla de manera inversa a como lo hacía el sujeto; y **No-effect**, en donde el avatar no variaba su tasa de habla en absoluto.

La experimentación mostró que en la condición 1 –**Entrainment vs Dis-entrainment**– hubo preferencias por **Dis-entrainment**; mientras que en la condición 2 –**Entrainment vs No-effect**– no hubo preferencias significativas. Una posible interpretación para la preferencia por **Dis-entrainment** podría ser que existe una suerte de contrapunto entre la tasa del habla del usuario realizando preguntas y el avatar respondiéndolas. Es decir, de alguna manera los dos actos de diálogo (pregunta y respuesta) podrían tener tasas de habla complementarias [20], lo cual sería medido como **dis-entrainment** según nuestra definición.

4.1. Aportes, limitaciones y trabajo futuro

Al hacer un balance sobre el trabajo realizado, notamos que el mayor aporte fue el desarrollo del juego y la evaluación de su utilidad para validar hipótesis sobre el efecto del entrainment prosódico. Si bien ya existía un desarrollo de un juego anterior –el original implementado en la tesis doctoral de Rivka Levitan [8]–, el hecho de que fuera un código multiplataforma permitió que el juego pueda ser adaptado para realizar experimentos en otros idiomas y con adaptación de la voz en más dimensiones, haciendo que se puedan extender los resultados de esta tesis. El código del juego quedó disponible para compartir (para conseguir una copia, contactar a los directores de la tesis).

Con el mismo juego se han desarrollado experimentos que permitieron escribir trabajos sobre el entrainment en sistemas de diálogo en distintos idiomas (en Inglés, Eslovaco y Español) extendiendo los resultados de esta tesis y su análisis [21].

Si repensamos el experimento realizado en comparación con las interacciones entre humanos, la mimetización que hacemos los humanos en nuestras conversaciones, es en múltiples dimensiones simultáneamente: el tono, el volumen, la tasa de habla, etc. Es por eso que al concluir este experimento, sospechamos que si la adaptación hubiese sido en múltiples dimensiones, como el tono, los resultados quizás habrían sido más contundentes. Por esto, como trabajo a futuro en esta misma línea de investigación nos parece pertinente que se investigue cómo reaccionan estos experimentos al adaptar el habla de estos SDS en más dimensiones prosódicas al mismo tiempo. En ese sentido, y con la utilización del mismo juego desarrollado para esta tesis ya hay un trabajo en preparación [22].

Es menester mencionar como limitación de esta tesis el funcionamiento del sintetizador de habla, ya que el utilizado (basado en `MaryTTS`) no fue exacto cuando se le solicitaba una variación muy pronunciada de la tasa de habla, es por eso que sería pertinente un trabajo a futuro que evaluara este mismo experimento con un TTS más exacto.

Otro aporte de esta tesis consiste en que los datos recolectados se suman a los de otros experimentos similares realizados en el marco del mismo proyecto de investigación. En su conjunto, estos datos posibilitarán analizar más aristas del fenómeno de entrainment. Por ejemplo, con los formularios de personalidad recolectados, podrá estudiarse el efecto de ciertas características de la personalidad de los sujetos en los resultados observados.

5. ANEXO I

En este anexo se muestran todos los formularios que debieron responder los sujetos que fueron expuestos al experimento. Si bien no todos fueron utilizados para interpretar resultados en este trabajo, los datos recolectados se suman a los de otros experimentos similares realizados en el marco del mismo proyecto de investigación. En su conjunto, estos datos posibilitarán analizar más aristas del fenómeno de entrainment

Las instrucciones que se le han entregado a los sujetos que llevaron a cabo el experimento, cumplen la función de inducir al sujeto a que juegue de acuerdo a la confianza que le provoque el avatar -consejera- por eso se le plantea que cada consejera utiliza un algoritmo diferente en la elección de su consejo. Así como también se lo mencionó verbalmente en cada sesión del experimento para afirmar el concepto.

Instrucciones para jugar Freir Churros con Ayudantes

El juego que vas a jugar se llama Freir Churros con Ayudantes. Es un juego de cartas para dos jugadores. Vos vas a ser uno de los jugadores y tu contrincante será la computadora. Una particularidad del juego es que vas a jugar al mismo a través de lo que te indiquen dos consejeras: Eugenia y Amanda.

La dinámica del juego es la siguiente. Inicialmente a cada jugador se le reparten siete cartas. Una vez hecho esto, vos tenés que elegir a alguna de tus dos consejeras y pedirle su consejo. Su consejo va a ser una indicación referida a qué carta tenés que pedirle a tu contrincante. Vos tenés la obligación de seguir el consejo. Sólo se te va a sugerir que pidas cartas de un número que vos tengas entre las tuyas (pudiendo este número ser J, Q, K y A). Una vez que pidas las cartas que tu consejera te recomendó, tu contrincante está obligado a darte todas las cartas que él tenga de ese número. Aquí pueden darse tres situaciones:

1. Tu contrincante te da sus cartas y vos pasás a tener los cuatro palos de ese número de carta (por ejemplo, pasás a tener las cuatro cartas con número “3”). En este caso ganás 10 puntos por cada carta que recibas de tu contrincante y 100 puntos más por haber formado un juego.
2. Tu contrincante te da sus cartas pero aun así no pasás a tener las cuatro cartas de ese número de distinto palo. En este caso ganás 10 puntos por cada carta que recibas y conservás las cartas.
3. Tu contrincante no tienen ninguna carta del número que le pediste. En este caso se te manda a “Freir Churros”. Este caso te obliga a levantar una carta del mazo y perdés 50 puntos

Luego de esto, tu contrincante va a devolver todas sus cartas restantes al mazo, las mismas se van a mezclar y se le volverá a repartir a él siete nuevas cartas.

Vos vas a interactuar con el sistema hablándole al mismo, te pedimos que por favor hables con el mismo de forma natural.

Cuestionario de Información personal

1. Sexo: Masculino Femenino
2. Edad: _____
3. ¿Qué mano usás para escribir? Izquierda: Derecha:

Idiomas:

4. Lugar de nacimiento: _____
5. ¿Cuál fue el primer idioma que aprendiste a hablar en forma fluida? _____
6. ¿Qué idioma hablaba tu madre en tu casa durante tu infancia? _____
7. ¿Qué idioma hablaba tu padre en tu casa durante tu infancia? _____
8. ¿Qué idioma hablaban tu madre y tu padre entre sí en tu casa durante tu infancia? _____
9. ¿Hablás algún otro idioma en forma fluída? Sí: No:

En caso afirmativo, listá los idiomas que hablás (incluído el español), especificando en cada caso si lo aprendiste en la escuela (ESC), o por escucharlo y usarlo regularmente en un entorno en el cual se hablaba ese idioma (ENT).

- _____ ESC: ENT : (desde la edad: _____)
- _____ ESC: ENT : (desde la edad: _____)
- _____ ESC: ENT : (desde la edad: _____)

Audición y Visión:

10. ¿Tenés audición normal? Sí: No: Corregida (audífono, etc.):
11. ¿Tenés visión normal? Sí: No: Corregida (anteojos, etc.):

Otros Antecedentes:

12. ¿Alguna vez tomaste clases de canto o música? Sí: No:
(Describir: _____)
13. ¿Cantás o tocás un instrumento frecuentemente? Sí: No:
(Describir: _____)
14. ¿Cursaste materias relacionadas a la Lingüística? Sí: No:
(Describir: _____)

Cuestionario de Preferencias

Por favor, respondé las siguientes preguntas:

¿Cuál consejera te gustó más? _____

¿Cuál de las voces te gustó más? _____

¿Cuál consejera dio mejores consejos? _____

Por favor, marcá con una cruz los adjetivos que usarías para describir la voz de {Amanda / Eugenia}.

Adjetivo	Amanda	Eugenia
Alta	<input type="checkbox"/>	<input type="checkbox"/>
Rápida	<input type="checkbox"/>	<input type="checkbox"/>
Suave	<input type="checkbox"/>	<input type="checkbox"/>
Lenta	<input type="checkbox"/>	<input type="checkbox"/>
Agradable	<input type="checkbox"/>	<input type="checkbox"/>
Natural	<input type="checkbox"/>	<input type="checkbox"/>
Molesta	<input type="checkbox"/>	<input type="checkbox"/>
Rara	<input type="checkbox"/>	<input type="checkbox"/>

Bibliografía

- [1] J. Edlund, J. Gustafson, M. Heldner, and A. Hjalmarsson, “Towards human-like spoken dialogue systems,” *Speech Communication*, vol. 50, no. 8-9, pp. 630–645, 2008.
- [2] N. Lubold, H. Pon-Barry, and E. Walker, “Naturalness and rapport in a pitch adaptive learning companion,” in *IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, 2015, pp. 103–110.
- [3] A. Papangelis, M. Kotti, and Y. Stylianou, “Predicting dialogue success, naturalness, and length with acoustic features,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 5010–5014.
- [4] S. Benus, A. Gravano, R. Levitan, S. Levitan, L. Willson, and J. Hirschberg, “Entrainment, dominance and alliance in Supreme Court hearings,” *Knowledge-Based Systems*, vol. in press, 2014.
- [5] S. E. Brennan and H. H. Clark, “Conceptual pacts and lexical choice in conversation,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 22, no. 6, p. 1482, 1996.
- [6] C. De Looze, S. Scherer, B. Vaughan, and N. Campbell, “Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction,” *Speech Communication*, vol. 58, pp. 11–34, 2014.
- [7] A. Gravano, S. Benus, R. Levitan, and J. Hirschberg, “Backward mimicry and forward influence in prosodic contour choice in Standard American English,” in *Interspeech*, 2015, pp. 1839–1843.
- [8] R. Levitan, “Acoustic-prosodic entrainment in human-human and human-computer dialogue,” Ph.D. dissertation, Columbia University, 2014.
- [9] R. C. Mayer, J. H. Davis, and F. D. Schoorman, “An integrative model of organizational trust,” *Academy of Management Review*, vol. 20, no. 3, pp. 709–734, 1995.
- [10] T. L. Chartrand and J. A. Bargh, “The chameleon effect: the perception–behavior link and social interaction.” *Journal of Personality and Social Psychology*, 1999.
- [11] A. Nenkova, A. Gravano, and J. Hirschberg, “High frequency word entrainment in spoken dialogue,” *Proceedings of ACL/HLT*, p. 169, 2008.
- [12] D. Reitter and J. Moore, “Alignment and task success in spoken dialogue,” *Journal of Memory and Language*, vol. 76, pp. 29–46, October 2014.
- [13] R. Levitan, A. Gravano, L. Willson, S. Benus, J. Hirschberg, and A. Nenkova, “Acoustic-prosodic entrainment and social behavior,” in *Proceedings of NAACL/HLT*, 2012.
- [14] C. Lee, A. Katsamanis, M. Black, B. Baucom, P. Georgiou, and S. Narayanan, “An analysis of PCA-based vocal entrainment measures in married couples’ affective spoken interactions,” in *Interspeech*, 2011.

-
- [15] J. Manson, G. Bryant, M. Gervais, and M. Kline, “Convergence of speech rate in conversation predicts cooperation,” *Evolution and Human Behavior*, vol. 34, p. 419–426, 2013.
- [16] C. Danescu-Niculescu-Mizil, L. Lee, B. Pang, and J. Kleinberg, “Echoes of power: language effects and power differences in social interaction,” in *Proc. of WWW*, 2012.
- [17] D. Huggins-Daines, M. Kumar, A. Chan, A. W. Black, M. Ravishankar, and A. I. Rudnicky, “Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices,” in *2006 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2006, pp. 185–188.
- [18] L. Violante, P. R. Zivic, and A. Gravano, “Improving speech synthesis quality by reducing pitch peaks in the source recordings,” in *Proceedings of HLT-NAACL*, 2013, pp. 502–506.
- [19] H. M. Torres and J. A. Gurlekian, “Automatic determination of phrase breaks for argentine spanish,” in *Speech Prosody*, 2004.
- [20] U. D. Reichel, N. Pörner, D. Nowack, and J. Cole, “Analysis and classification of cooperative and competitive dialogs,” in *Interspeech*, 2015, pp. 3056–3060.
- [21] R. Levitan, Š. Beňuš, R. H. Gálvez, A. Gravano, F. Savoretti, M. Trnka, A. Weise, and J. Hirschberg, “Implementing acoustic-prosodic entrainment in a conversational avatar,” *Interspeech 2016*, 2016.
- [22] R. H. Gálvez, “An empirical study on how several policies of acoustic-prosodic entrainment affect the perceived trustworthiness of conversational avatars,” *en preparación*, 2018.