



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Estudio del efecto de la mimetización acústico-prosódica en la interacción con un sistema de diálogo hablado

Buenos Aires

Noviembre 2017

Tesis de Licenciatura en Ciencias de la Computación

Directores: Agustín Gravano y Ramiro H. Gálvez

Gauder María Lara	027/10	marialaraa@gmail.com
Reartes Marisol Elizabeth	422/10	mreartes5@gmail.com



Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2160 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (54 11) 4576-3359

<http://www.fcen.uba.ar>

Estudio del efecto de la mimetización acústico-prosódica en la interacción con un sistema de diálogo hablado

La tecnología lleva a los seres humanos a interactuar constantemente con computadoras y smartphones que tienen incorporados sistemas de diálogo hablado. Sin embargo, aún no se pudo lograr un sistema capaz de comunicarse con total naturalidad. El objetivo de esta tesis es estudiar los efectos que tiene sobre la naturalidad de las interacciones la mimetización acústico-prosódica (la propensión a adaptar características del habla a la del interlocutor). Específicamente se ve qué efecto produce sobre los sujetos el entrainment y disentrainment acústico-prosódico mediante sincronía aplicados a los atributos prosódicos del tono, intensidad y tasa del habla. Se desea conocer qué política de mimetización se asocia a un mayor nivel de competencia o habilidad. Para ello se desarrolló un sistema de diálogo hablado basado en el popular juego “¿Quién es quién?”, con una estrategia de Mago de Oz ante la falta de reconocedores automáticos que aseguren una transcripción exacta de lo dicho por el usuario. El juego consta de tres rondas y en cada una de ellas el sujeto juega contra dos avatares (cada uno implementa una política de mimetización distinta). Luego, el sujeto debe elegir contra quién quiere jugar una partida final bonificada. En el análisis de los resultados, se observó una tendencia a considerar como más competentes a los avatares con habla con disentrainment.

Palabras claves: entrainment, sistema de diálogo hablado, mago de oz, competente, confianza

Índice

1. Introducción	1
2. Métodos	3
2.1. En qué consiste el juego	3
2.2. Estructura del juego	3
2.2.1. Dinámica	3
2.2.2. Avatares	3
2.2.3. Modos de juego	4
2.2.4. Rondas	5
2.2.4.1. Ronda de práctica	6
2.2.4.2. Rondas principales	7
2.2.4.3. Ronda de partidas finales	8
2.3. Interfaz del juego	8
2.3.1. Interfaz del sujeto	8
2.3.2. Escenas	9
2.3.3. Propiedades	10
2.3.4. Tableros y cartas	11
2.3.5. Botón push-to-talk	13
2.3.6. Tablas de resultados	13
2.3.7. Instrucciones dirigidas al sujeto	14
2.4. Decisiones de diseño del juego	14
2.4.1. Balanceo del tablero	14
2.4.2. Generación de preguntas	15
2.4.2.1. Backtracking	15
2.4.2.2. Preguntas predefinidas y reglas según condiciones	15
2.4.3. Manipulación	16
2.4.3.1. Cantidad de preguntas necesarias	16
2.4.3.2. Resultados de las partidas	16
2.4.4. Recuperación ante fallas	17

2.5. Mago de Oz	17
2.5.1. Interfaz	18
2.5.2. Comunicación con el juego	19
2.6. Procesamiento de los atributos acústicos-prosódicos	20
2.6.1. Procesamiento del audio	20
2.6.1.1. Reconocimiento del habla	20
2.6.1.2. Síntesis del habla	21
2.6.1.3. Características del habla del sujeto	22
2.6.2. Speech history	23
2.6.3. Actos de diálogo	23
3. Resultados	25
3.1. Recolección de datos	25
3.2. Análisis según la política de mimetización	27
3.3. Análisis de subgrupos	27
3.4. Análisis según el resultado de las últimas partidas de cada ronda	30
3.5. Análisis de potencia	31
4. Conclusiones y trabajo futuro	32
A. Secuencias de las rondas	34
B. Tableros de las rondas	35
C. Experimento	37
C.1. Cheat Sheet	37
C.2. Texto introductorio	38
C.3. Consentimiento	39
C.4. Formulario de personalidad	41
C.5. Instrucciones del juego	42
C.6. Formulario de formación	44
C.7. Recibo	45
C.8. Detalles del experimento	46

D. Resultados complementarios

47

1. Introducción

Hoy en día la tecnología lleva a los seres humanos a interactuar constantemente con computadoras y smartphones. Muchos de ellos tienen incorporados **sistemas de diálogo hablado** (SDH) para facilitar el uso de los dispositivos. Un SDH es un sistema informático capaz de conversar con un humano oralmente. El constante aumento de la utilización de estos aparatos contribuyó al crecimiento de las tecnologías, principalmente a los sistemas de reconocimiento automáticos de habla (ASR) y los sistemas de texto a habla (TTS). Sin embargo, a pesar de los grandes avances de la informática, aún no se pudo lograr un sistema capaz de comunicarse con una persona tal como si lo hiciera con otra. En otras palabras, aún faltan avances que permitan total naturalidad en la comunicación. En un sistema de diálogo hablado todavía no son claras qué propiedades debe cumplir la computadora para lograr la naturalidad en la conversación.

Un fenómeno que ocurre en el diálogo humano-humano es el **entrainment**, también es conocido como *mimetización*. El entrainment es definido como la tendencia de una persona a adaptarse o imitar el comportamiento del habla de su interlocutor. Este fenómeno también es conocido como coordinación, adaptación o alineación. Los humanos presentan entrainment no sólo en el habla sino también en otros aspectos, como en gestos [Chartrand and Bargh, 1999], rasgos acústicos [Natale, 1975], fonéticos [Pardo, 2006], léxicos [Brennan and Clark, 1996] y sintácticos [Reitter and Moore, 2007].

Se pueden diferenciar tres tipos de entrainment acústico-prosódico [Levitan and Hirschberg, 2011]. En primer lugar, podemos distinguir a la *proximidad*, que hace referencia a la similitud de un atributo en toda la conversación. Por otro lado, la *convergencia* es cuando los valores de los atributos se van acercando cada vez más durante un diálogo. Por último, diferenciamos la *sincronía* que implica que los hablantes mantienen una coordinación de su habla a lo largo de toda la charla. En este estudio se analizó el comportamiento de los sujetos con respecto al entrainment mediante sincronización.

La evidencia de entrainment en casi todos los niveles de comunicación sugiere su importancia en los sistemas de diálogo hablado. En base a estudios previos que constatan que en una conversación entre seres humanos se produce entrainment, se trató de replicar este fenómeno en una computadora para ver qué efectos se producen sobre los sujetos.

En este trabajo analizamos el entrainment a nivel acústico-prosódico. Puntualmente nos enfocamos en el entrainment generado a partir de tres atributos prosódicos¹. Estos son: **intensidad** (intensity), **tono** (pitch) y **tasa del habla** (speech rate). Se manipularon estos valores para poder generar una sincronía entre las características prosódicas del habla del sistema y la de los sujetos.

Además, en este proyecto las conversaciones involucran principalmente dos actos de diálogo (tipo específico de acción que se puede dar en las conversaciones) distintos. Por un lado, en las conversaciones ocurren preguntas que admitan como respuesta sí o no y, por otro, respuestas de tipo sí o no combinadas con una declaración [Jurafsky et al., 1998].

Otro fenómeno que ocurre en los diálogos entre humanos es el **disentrainment**. Siguiendo la terminología de Pérez et al. [2016], llamamos disentrainment a un fenómeno opuesto al entrainment, cuando un hablante sube el nivel de algún atributo, su interlocutor lo baja y viceversa. En otras palabras, los atributos prosódicos de la voz del interlocutor se alejan de los del hablante.

A partir del estudio de Pérez et al. [2016] se observó una correlación entre el fenómeno de disentrainment y un número positivo de características sociales de las conversaciones. Se decidió ver qué ocurre al introducir entrainment y disentrainment a nivel de actos de diálogo. En este trabajo, se aplica las políticas de mimetización en dos actos de diálogo distintos. Por un lado, entre las preguntas del sujeto y las de la computadora y por el otro, entre las respuestas de las mismas partes.

Según el paper BLA, consideramos confianza de un sujeto hacia otro cuando uno de ellos delega una tarea que implica un riesgo para él. La confianza se puede dar en tres dimensiones distintas:

¹La prosodia es una rama de la lingüística que analiza y representa formalmente aquellos elementos de la expresión oral tales como el acento, los tonos y la entonación.

habilidad, benevolencia e integridad. La confianza en *habilidad* hace referencia a la capacidad del sujeto en resolver la tarea delegada. Por otro lado, *benevolencia* es el beneficio personal que pueda obtener el receptor de la confianza en resolver la tarea. Por último, *integridad* está relacionado con que los principios del dador de la confianza coincidan con los del sujeto al que se le delega la tarea. En esta tesis se hace énfasis en la confianza en *habilidad*.

La motivación de esta tesis es mejorar la comunicación entre las personas y la computadora para generar un diálogo más natural y ameno. Por lo tanto, el interrogante a responder de esta tesis es: **¿Existe una relación causal entre la mimetización acústico-prosódica y el grado de confianza generado en el hablante?** El objetivo primario es el desarrollo de un SDH para testear la hipótesis planteada. Otro objetivo de esta tesis es estudiar la factibilidad del SDH implementado para responder el interrogante central, a nivel de actos de diálogo.

Para investigar esta relación se llevaron a cabo experimentos propios con un sistema de diálogo hablado. A diferencia de otros trabajos de investigación relacionados al fenómeno mencionado no se usó un corpus de datos ya existente. El SDH consiste en una implementación del popular juego “**¿Quién es quién?**”, desarrollado específicamente para esta tesis. El juego es muy similar al original pero con algunas modificaciones para lograr los efectos deseados sobre los sujetos.

En el juego se da una interacción de preguntas y respuestas entre un sujeto y una computadora. Las preguntas y respuestas son hechas tanto por los sujetos como por la computadora. Es decir, la conversación es simétrica, ambas partes no hacen siempre una única acción durante todo el juego, sino que se van alternando. Todo esto hace más interesante al sistema, por ende, también a los resultados que se pueden obtener. Además, es novedoso con respecto a otros estudios realizados con el mismo fin. El juego presenta diferentes modos que van alternando ciertos factores definidos y que servirán para el análisis. Los mismos serán explicados en más detalle.

La falta de reconocedores automáticos que aseguren una transcripción exacta de lo que dicen los sujetos llevó a implementar una estrategia diferente. La misma consiste en tener una persona que interprete lo que el sujeto va diciendo. Esta persona es llamada **Mago de Oz** (Wizard of Oz) [Ter Maat et al., 2010]. Éste permite flexibilidad en el habla de los sujetos y mayor precisión en la traducción.

En otros trabajos se encuentran, a partir del corpus, correlaciones entre distintas variables. Lo que se intenta buscar a partir de este proyecto no es una correlación de los hechos sino una causalidad. Sólo una variable de las que podrían influir en el juego es diferente en las distintas variantes del juego y controlamos todas las otras variables posibles. Entonces, al obtener los resultados, la única causa posible es la variable que cambia. Se quiere ver que la variable sea diferente del azar. En este proyecto lo único que se varía es la política de mimetización, ya sea entrainment o disenrainment. Por lo tanto, cualquier efecto observado sólo puede deberse a la variación de dicha variable.

Este informe se encuentra estructurado de la siguiente manera. En la Sección 2 se presentan los métodos, se describen tanto las decisiones tomadas referidas al diseño del experimento y al de la interfaz. La Sección 3 presenta los resultados de los experimentos realizados. Por último, en la Sección 4 se describen las conclusiones obtenidas y se proponen mejoras para trabajos futuros.

2. Métodos

En este capítulo, presentamos la metodología empleada en esta tesis. Describimos el sistema de diálogo hablado desarrollado: las reglas del juego, su estructura, las interfaces de usuario, la estrategia del Mago de Oz y las componentes de procesamiento del habla.

2.1. En qué consiste el juego

El juego “*¿Quién es Quién?*” es un juego de mesa que involucra a dos personas. Cada sujeto tiene un tablero con 32 cartas y una aparte, que es la que debe adivinar su oponente. Cada carta contiene un personaje. Ambos tableros tienen las mismas cartas ordenadas de manera diferente y la carta a adivinar se encuentra entre ellas.

A partir de preguntas cerradas que van haciendo los jugadores en distintos turnos, van descartando cartas de su propio tablero según las respuestas. Las preguntas deben ser por alguna característica de los personajes de las cartas del tablero, y deben ser formuladas de tal forma que las respuestas sean por sí o por no. Un ejemplo de pregunta es: “¿Tiene ojos celestes?”.

Cuando a uno de los dos le queda una única carta, entonces el oponente debe arriesgarse por alguna de las que no haya bajado en su tablero. Si este último acierta, entonces gana la partida. Caso contrario, el ganador es el jugador al que le había quedado una única carta en su tablero.

2.2. Estructura del juego

En esta sección se describe cómo está organizado el juego “*¿Quién es quién?*” implementado y los avatares involucrados. Además, se explican las distintas variantes del juego.

2.2.1. Dinámica

El juego consta de tres rondas y en cada una el sujeto juega contra dos oponentes/avatares diferentes para poder ver el nivel de cada uno. Al terminar cada ronda debe elegir contra quién quiere jugar la partida final de cada una. Por partida final ganada se le da al sujeto un bono extra monetario. Éste nos permite estimular al sujeto para que se concentre en determinar cuál de los dos avatares es el más inteligente y, así, elegir al contrario.

Se tomó la decisión de que el sujeto sea el que siempre empiece una partida. De esta forma, siempre hace la primer pregunta y, así, se evita generarle confusiones sobre la dinámica de las partidas.

2.2.2. Avatares

En el juego hay **avatares**, que son los oponentes de los sujetos. Por cada ronda hay dos y en la misma uno sigue una política de entrainment acústico-prosódico y el otro de disentrainment. El análisis de este proyecto se basa en las elecciones del sujeto con respecto al habla de los avatares. Por este motivo, es muy importante acotar la cantidad de factores distintos que hagan que un sujeto decida jugar contra un avatar y no el otro.

Para reducir los factores que pueden influenciar, se dibujaron los avatares con los mismos rasgos, cambiando sólo el color de pelo y de la ropa. En la Figura 1 vemos los seis avatares distintos del juego. Natalia y Florencia corresponden a la ronda 1, Mariana y Susana a la ronda 2 y, por último, Sandra y Marina juegan en la ronda 3.

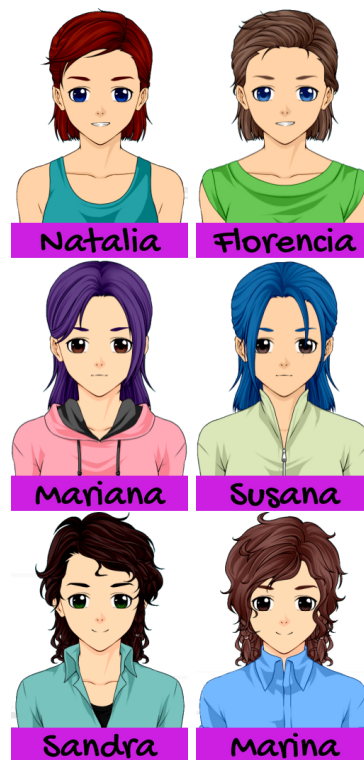


Figura 1: Imágenes de los avatares agrupados por ronda

2.2.3. Modos de juego

En el juego hay distintos tipos de factores que podrían influir en la elección del avatar. Por ejemplo, el sujeto podría tener una preferencia por un avatar basándose en su apariencia, como el color de pelo o de ojos. Otro motivo, podría ser el peso que un sujeto puede darle a lo que escucha primero o último: podría dar prioridad a una voz con entrainment sólo por ser la primera que escucha.

Por estas razones, se armaron cuatro **modos de juego** distintos variando estos factores. En cada uno se modifica la política de mimetización con la que va a hablar el avatar durante su ronda. Por otro lado, se altera el orden de participación de los avatares. De intercalar todas estas variantes, se obtiene la tabla de la Figura 2.

<div>1</div> <div>Ronda 1 Comienza: Oponente 1 Oponente 1 (Natalia): Entrainment Oponente 2 (Florescia): Disentrainment</div> <div>Ronda 2 Comienza: Oponente 4 Oponente 3 (Mariana): Entrainment Oponente 4 (Susana): Disentrainment</div> <div>Ronda 3 Comienza: Oponente 5 Oponente 5 (Sandra): Entrainment Oponente 6 (Marina): Disentrainment</div>	<div>2</div> <div>Ronda 1 Comienza: Oponente 2 Oponente 1 (Natalia): Entrainment Oponente 2 (Florescia): Disentrainment</div> <div>Ronda 2 Comienza: Oponente 3 Oponente 3 (Mariana): Entrainment Oponente 4 (Susana): Disentrainment</div> <div>Ronda 3 Comienza: Oponente 6 Oponente 5 (Sandra): Entrainment Oponente 6 (Marina): Disentrainment</div>
<div>4</div> <div>Ronda 1 Comienza: Oponente 2 Oponente 1 (Natalia): Disentrainment Oponente 2 (Florescia): Entrainment</div> <div>Ronda 2 Comienza: Oponente 3 Oponente 3 (Mariana): Disentrainment Oponente 4 (Susana): Entrainment</div> <div>Ronda 3 Comienza: Oponente 6 Oponente 5 (Sandra): Disentrainment Oponente 6 (Marina): Entrainment</div>	<div>3</div> <div>Ronda 1 Comienza: Oponente 1 Oponente 1 (Natalia): Disentrainment Oponente 2 (Florescia): Entrainment</div> <div>Ronda 2 Comienza: Oponente 4 Oponente 3 (Mariana): Disentrainment Oponente 4 (Susana): Entrainment</div> <div>Ronda 3 Comienza: Oponente 5 Oponente 5 (Sandra): Disentrainment Oponente 6 (Marina): Entrainment</div>

Figura 2: Tabla comparativa de las diferencias entre los distintos modos de juego

Para entender mejor cómo se llevan a cabo los modos de juego veamos un ejemplo: El sujeto está jugando en el modo de juego 3. En la ronda 1 le toca jugar la primera partida contra Natalia, que adopta la política de disentrainment. La siguiente partida tiene como avatar a Florescia, un avatar con política de entrainment. Luego, se repite la misma secuencia. Finalmente, el sujeto indica contra cuál de los dos avatares quiere jugar la partida final. Por otro lado, si el modo de juego es el número 2, entonces le toca jugar la primer partida contra Florescia que tiene política de disentrainment. Luego, juega contra Natalia con política entrainment.

Además, se realizaron los experimentos de manera tal que en cada grupo se lleven a cabo con la misma cantidad de hombres y mujeres. De esta manera, nos aseguramos que en caso de querer analizar los resultados distinguiendo el género, haya la misma cantidad de resultados para cada uno.

2.2.4. Rondas

El juego se divide en cinco rondas. Se inicia el juego con una ronda de práctica, le siguen tres rondas principales y, finalmente, la ronda de partidas finales. Esta secuencia se ilustra en la Figura 3.



Figura 3: Secuencia de las rondas que el sujeto irá jugando

2.2.4.1. Ronda de práctica

En la ronda de práctica el sujeto juega una única partida contra un único avatar (ver Figura 4).

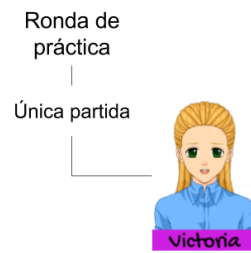


Figura 4: Ronda de práctica para todos los modos de juego

El objetivo de esta partida es familiarizar al sujeto con la dinámica del juego. Además, es un buen momento para que evacue todas sus dudas antes de comenzar con el experimento. Por otro lado, se utiliza esta partida para obtener los valores base para los audios de los avatares de la primera partida de la primera ronda (como se explica en la Sección 2.6.3).

Otro fin de esta ronda es que el Wizard verifique durante la partida lo siguiente:

- el uso correcto de las propiedades;
- el manejo del botón de audio (si deja mucho tiempo en silencio antes de soltarlo o al comienzo, si se corta en el medio, por ejemplo);
- la velocidad con la que habla (que no hable ni muy rápido ni muy despacio);
- la claridad con la que habla el sujeto;
- la longitud de las preguntas y respuestas (que siga las estructuras indicadas);
- la distancia entre el micrófono y la boca del sujeto (la respiración podría afectar los audios);
- si el sujeto escucha bien;
- si toca/mueve el micrófono;
- otras observaciones.

Luego, al terminar esa partida al sujeto le aparece un cartel indicando que avise que terminó la misma. En caso que el Mago de Oz haya anotado alguna de las observaciones anteriores se le dice al sujeto cómo corregirlas. A continuación, se le explica cómo es la dinámica del juego a partir de ese momento.

La partida de práctica tiene como avatar a Victoria (ver Figura 5). Ésta se presenta como quien va a ir explicando ciertos detalles del juego. Se utilizan globos explicativos para ir comentando algunos componentes del juego y las acciones que se deben ir realizando. Victoria utiliza voz neutra, es decir, son constantes sus valores de tono, intensidad y tasa a lo largo de la partida.

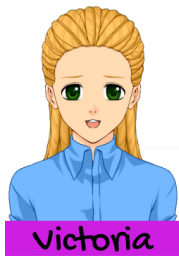


Figura 5: Imagen del avatar de la ronda de práctica

2.2.4.2. Rondas principales

Las rondas principales son aquellas en las que los sujetos deben elegir un avatar para jugar la partida final de la misma. La cantidad de rondas principales es impar. Así, para cada sujeto tenemos una política de mimetización ganadora (entrainment o disenitainment) y evitamos los empates. Esto es necesario para uno de los análisis estadísticos a realizar.

Además, es razonable que el experimento no dure mucho más que una hora para evitar el cansancio del sujeto. Por lo tanto, habiendo estimado cuánto lleva completar una ronda del juego, se determinó que el mismo dure tres rondas. Se llegó a este número ya que una ronda sería poco para obtener resultados y cinco harían muy extenso el experimento.

La Figura 6 ilustra una posible configuración de dos modos de juego. El sujeto juega tres rondas de cuatro partidas cada una. En la imagen se puede apreciar la secuencia de partidas con el avatar de cada una y su resultado.

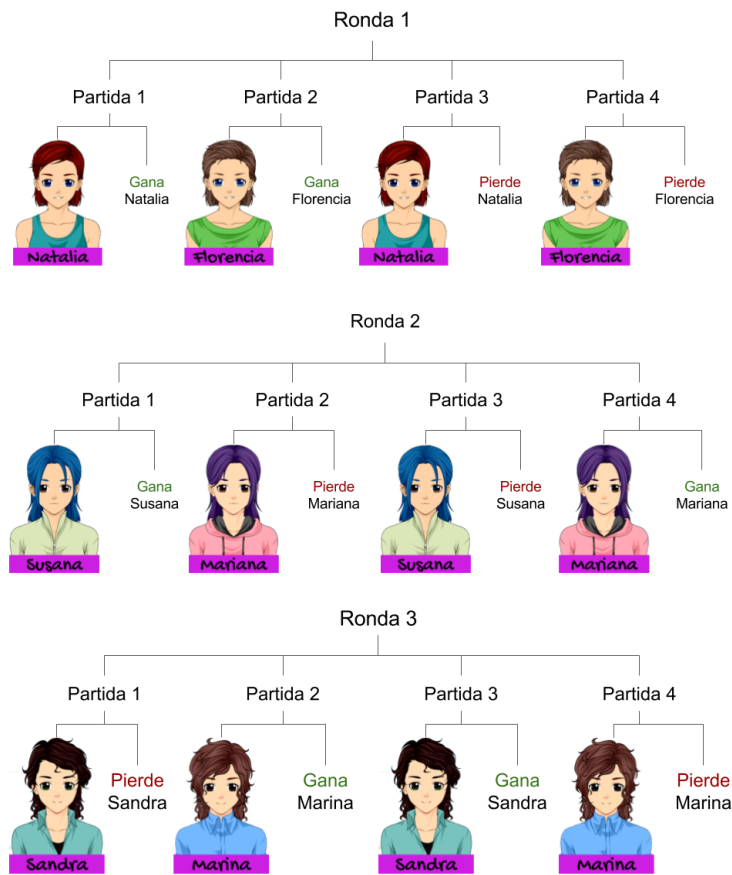


Figura 6: Configuración de las rondas 1, 2 y 3 para los modos de juego 1 y 3

Cabe destacar que se fuerza a que el sujeto gane y pierda una vez contra cada avatar en cada ronda. Es importante que el sujeto no condicione su elección final en base a los resultados de las partidas de la ronda.

2.2.4.3. Ronda de partidas finales

Luego de que el sujeto juega las cuatro partidas de una ronda contra los dos avatares correspondientes debe elegir contra cuál jugar la partida final. Estas partidas se juegan en la ronda de partidas finales.

Esta ronda se juega una vez finalizadas las tres rondas principales. Se quiere evitar cualquier efecto que puedan tener las partidas finales sobre las posteriores elecciones de avatares. Por ejemplo, podría ser que el resultado de la quinta partida de la primera ronda influyera en la decisión de qué avatar elige en la siguiente ronda.

2.3. Interfaz del juego

En esta sección describimos las diferentes partes que componen la interfaz del juego, es decir, la que ve el sujeto.

2.3.1. Interfaz del sujeto

Describimos a continuación las distintas partes del tablero y su funcionalidad. La Figura 7 muestra las partes del tablero numeradas de 1 a 6.

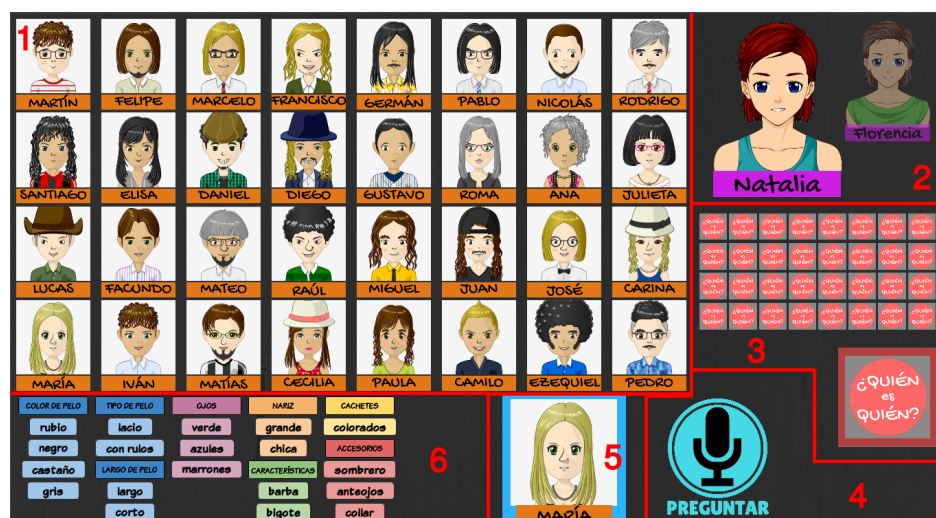


Figura 7: Interfaz del sujeto al comenzar la primera partida de la ronda 1

1. Tablero del sujeto.

Lo utiliza como guía para ir descartando cartas a medida que hace preguntas.

2. Avatares.

Se muestran los dos avatares contra los que el sujeto juega a lo largo de la ronda. Se destaca al avatar actual.

3. Tablero y carta del avatar.

De esta manera, el sujeto puede saber el avance del avatar. La carta del avatar es la que debe adivinar el sujeto. Ésta se muestra al terminar la partida.

4. Botón de audio con la etiqueta correspondiente.
5. Carta que debe adivinar el avatar.
6. Propiedades que el sujeto puede preguntar en esa ronda.

2.3.2. Escenas

A continuación, se describen las escenas más destacadas del juego, mostrando su imagen y el contexto donde ocurre.

Una de las escenas más comunes ocurre luego de que el avatar responde la pregunta hecha por el sujeto. Se muestra un cartel indicando que se van a bajar las cartas que cumplen con cierta propiedad (ver Figura 8).



Figura 8: Interfaz del sujeto cuando el avatar responde la pregunta que se le hizo

Otra escena recurrente se da cuando la partida finaliza. Se muestra un cartel indicando si ganó o perdió y la tabla con los resultados hasta el momento para esa ronda. Se puede ver un ejemplo de cada caso en la Figura 9.



Figura 9: Posibles interfaces del sujeto al terminar una partida

Es importante destacar al sujeto el comienzo de una nueva ronda, así como también, los dos avatares contra los que juega en la misma. Se pueden observar escenas de ejemplo en la Figura 10.

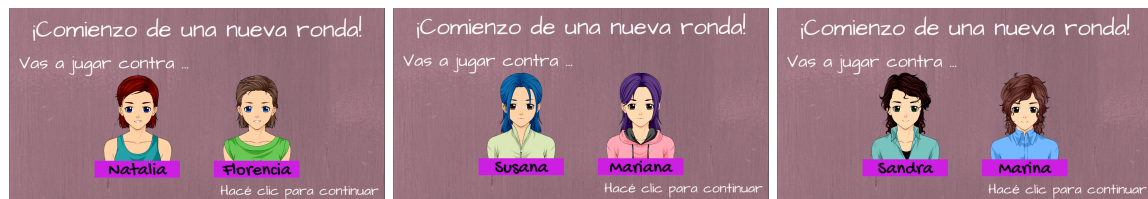


Figura 10: Escenas del juego que se muestra al comenzar cada una de las tres rondas

Consideramos fundamental destacarle al sujeto el avatar contra el que le toca jugar la nueva partida. Por ello, se resalta al mismo en una escena antes de comenzar la partida. Se puede ver un ejemplo en la Figura 11.

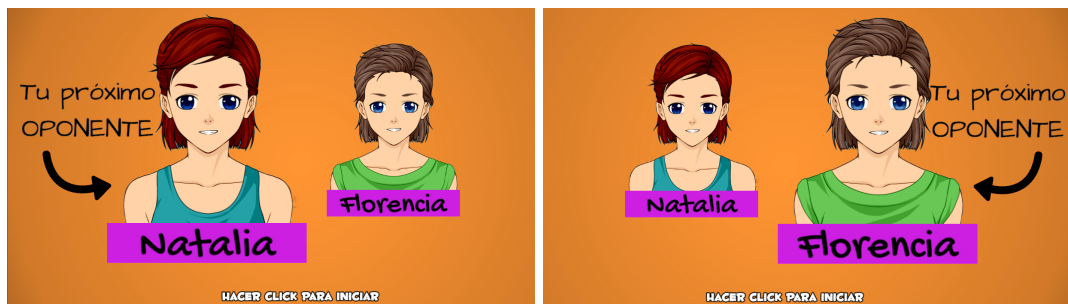


Figura 11: Escenas del juego que muestran el próximo avatar de la partida que comenzará

La última escena a distinguir es la que ocurre luego de terminar una ronda. El sujeto debe seleccionar oralmente contra qué avatar quiere jugar la partida final. Una vez dicho, se resalta el avatar que haya elegido. Se ven ejemplos de ambas escenas en la Figura 12.



Figura 12: Escenas del juego al momento de elegir un avatar para la partida final de la ronda 1

2.3.3. Propiedades

En el juego de mesa cada sujeto puede preguntar por cualquier propiedad (característica) que se vea en la imagen de la carta. Por otro lado, en el juego desarrollado para esta tesis se tomó la decisión de acotar las posibilidades de propiedades que se pueden preguntar. Esta decisión está influenciada por la funcionalidad del Mago de Oz que se explica en más detalle en la Sección 2.5.1.

Por este motivo, en una parte de la pantalla de la partida se muestra al sujeto el listado de las propiedades que puede consultar. La idea es que tenga siempre visibles las propiedades permitidas y no haga preguntas inválidas, en lo posible. Una propiedad válida en la primera ronda es *cachetes colorados*. Un ejemplo de pregunta correcta es: “¿El personaje de la carta que tengo que adivinar tiene los cachetes colorados?”.

El juego tiene un conjunto de propiedades diferente para cada ronda. En la primera y tercera ronda las cartas son personas, mientras que en la segunda ronda las cartas contienen figuras, como círculos y cuadrados.

En las reglas del juego de mesa no se puede preguntar por el género del personaje de la carta en el primer turno. En el experimento, para cada ronda se definió que haya una propiedad que recién se pueda preguntar a partir del cuarto turno. En el caso de la primera y tercera ronda es el género del personaje, mientras que en la segunda es el tipo de figura (cuadrado o círculo). Esto se puede ver en Figura 13). En la sección de propiedades de la interfaz no se muestran las imágenes con los nombres de estas propiedades hasta el cuarto turno.



Figura 13: Cambio de estado de la sección de la interfaz que muestra las propiedades que el sujeto puede preguntar para las tres rondas

2.3.4. Tableros y cartas

Un tablero está compuesto por 32 cartas. Para cada ronda principal se diseñó un conjunto de cartas distintas que constituyen el tablero correspondiente.

Se definió que las propiedades entre los tres tableros sean equivalentes. Por ejemplo, en la primera ronda la propiedad de color de pelo (con opciones: negro, marrón, rubio y gris) se corresponde con la característica de cantidad de figuras en la carta (con opciones: una, dos, tres y cuatro) de la ronda dos. Mientras que en la ronda tres se corresponde con el color de pelo (con opciones: naranja, azul, violeta y verde). La Tabla 1 muestra la correspondencia entre las propiedades de cada ronda.

Ronda 1	Ronda 2	Ronda 3
<i>Color de pelo</i>	<i>Cantidad de figuras</i>	<i>Color de pelo</i>
Rubio	Uno	Violeta
Negro	Dos	Azul
Marrón	Tres	Naranja
Gris	Cuatro	Verde
<i>Largo del pelo</i>	<i>Color del borde de la figura</i>	<i>Tamaño de la nariz</i>
Corto	Negro	Grande
Largo	Blanco	Chica
<i>Tamaño de la nariz</i>	<i>Color del fondo de la figura</i>	<i>Tamaño de la boca</i>
Grande	Azul	Grande
Chica	Rosa	Chica
<i>Tipo de pelo</i>	<i>Grosor del borde de la figura</i>	<i>Largo del pelo</i>
Lacio	Finito	Corto
Rulos	Grueso	Largo
<i>¿Tiene cachetes colorados?</i>	<i>Textura de la figura</i>	<i>¿Tiene pecas?</i>
Sí	Rayado	Sí
No	Liso	No
<i>Color de ojos</i>	<i>Color de fondo de la carta</i>	<i>Color de ojos</i>
Verdes	Verde	Verdes
Azules	Rojo	Azules
Marrón	Amarillo	Marrón
<i>Género</i>	<i>Tipo de figura</i>	<i>Género</i>
Masculino	Cuadrado	Masculino
Femenino	Círculos	Femenino
<i>Accesorio</i>	<i>Figuras agregadas</i>	<i>Accesorio</i>
Anteojos	Pica	Sombrero
Sombrero	Trébol	Anteojos
Ninguno	Ninguna	Ninguno
<i>Características especial</i>	<i>Figura especial</i>	<i>Características especial</i>
Barba	Estrella	Corbata
Bigote	Corazón	Moño
Collar	Flor	Collar

Tabla 1: Listado de las propiedades de las distintas rondas, con su equivalente en las demás

Además, cada ronda cuenta con un tablero diferente en propiedades pero igual en balanceo de las mismas. Los tableros tienen cartas distintas para evitar que el sujeto se familiarice con las cartas y pueda encontrar patrones en ellas. De esta manera, se intenta reducir la presencia de factores que puedan perjudicar el experimento. Se pueden ver los tableros en el Apéndice B. Otro motivo para establecer que cada ronda tenga un tipo de tablero propio es para que al sujeto no le sea tedioso jugar tanto tiempo con un mismo tablero. En la primera² y tercera³ ronda se utilizan dibujos de personas, pero en la segunda, se muestran figuras⁴. Se tomó la resolución de que las cartas del segundo tablero sean así para que no resulte monótono siendo siempre caras. Por la misma razón las caras de la primera y última ronda son diferentes, así como también algunas de las propiedades por las que se puede preguntar.

Asimismo, si una carta tiene una propiedad en una ronda, en otra va a tener una equivalente. Esto se propaga a los conjuntos de propiedades de las cartas, por lo que las mismas son equivalentes en todas las rondas. Por ejemplo, veamos el caso de las cartas de la Figura 14. La carta de *Ezequiel* de la ronda 1 es equivalente a la carta *Figura C6* de la ronda 2 y a la carta *Federico* de la ronda 3. Por

²Diseñadas usando el generador de www.avachara.com/avatar/

³Diseñadas utilizando el generador de www.faceyourmanga.com

⁴Diseñadas usando el programa *Draw* de *Libre Office* (www.libreoffice.org)

ejemplo, Ezequiel tiene la nariz chica, la Figura C6 de la ronda 2 tiene el fondo de las figuras rosa y Federico tiene su boca chica. Esta correspondencia entre propiedades puede ser vista en la Tabla 1.

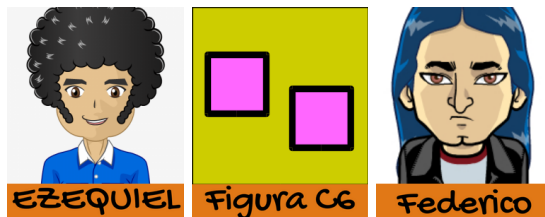


Figura 14: Comparación de tres cartas, siendo la primera de la ronda 1, la segunda de la ronda 2 y la tercera de la ronda 3

Con respecto a las cartas que deben adivinar ambos sujetos, se seleccionaron cartas equivalentes para cada ronda.

2.3.5. Botón push-to-talk

El sujeto se comunica a lo largo del experimento con el avatar a través de un sistema de diálogo hablado. Todas las preguntas, respuestas y selecciones debe hacerlas oralmente. Para esto se incluyó en el juego un botón de tipo *push-to-talk*.

Es un botón al estilo *Whatsapp*. El sujeto aprieta el botón, habla, y cuando finaliza lo suelta. Se eligió este tipo de botón para que el sujeto interactúe ya que se encuentra en muchas aplicaciones populares de hoy en día y las personas ya están familiarizadas con su funcionalidad.



Figura 15: Botón *push-to-talk*

Además, al oprimir este botón comienza la grabación del audio, y al soltarlo finaliza la misma. Por lo tanto, permite conocer con precisión el comienzo y final del habla en la señal de audio.

Debajo de la imagen del botón audio se agregaron etiquetas con los nombres de las diferentes acciones que se pueden realizar a la hora de preguntar/responder. El nombre de la etiqueta que se muestra es la siguiente acción que el sujeto debe realizar. Se le deja al sujeto durante el experimento una *cheat sheet* que tiene una tabla explicativa con la descripción de la acción a realizar para cada etiqueta. Se puede ver en detalle en el Apéndice C.1.

Durante el experimento surgió una problemática con respecto a la funcionalidad del botón. Si el sujeto, luego de mantenerlo apretado, lo suelta con la flecha del mouse fuera de la imagen (Figura 15), se genera un error en la aplicación. Éste es uno de los casos donde se aprovecha lo explicado en la Sección 2.4.4. Se optó por explicarle al usuario como usar el botón correctamente. Luego de eso, no volvió a ocurrir el error.

2.3.6. Tablas de resultados

Antes de comenzar la etapa de recolección de datos, se llevaron a cabo algunas pruebas piloto para asegurar el correcto funcionamiento de la aplicación. En una de ellas el sujeto nos indicó que había elegido al avatar para jugar la partida final según si había ganado o perdido contra el avatar de la cuarta partida. Si había ganado contra el último, entonces lo elegía; si no, optaba por el otro.

Para poder controlar esta situación, se agregaron tablas que le muestran al sujeto la cantidad de

partidas que ganó o perdió a medida que avanza en cada una. De esta manera, el sujeto es consciente de que, al terminar la ronda, ganó y perdió la misma cantidad de veces contra cada avatar. Ejemplos de estas tablas pueden ser vistos en la Figura 9.

2.3.7. Instrucciones dirigidas al sujeto

Cuando el sujeto pregunta por una propiedad no permitida o el audio no es claro se le indica que reformule o repita la oración. Inicialmente, ésto se implementó de manera tal que el avatar sea el encargado de indicar la acción al sujeto de manera hablada. La desventaja de esta implementación es que, si hay que dar instrucciones muchas veces en una misma partida, se desbalancea la cantidad de preguntas y respuestas con respecto a otras partidas.

Para que la conversación entre el avatar y el sujeto sea lo más controlable posible, se optó por mostrar este tipo de instrucciones en carteles al lado del botón de audio de la interfaz del sujeto.

2.4. Decisiones de diseño del juego

A lo largo de esta sección, se describen decisiones tomadas para lograr los objetivos planteados para esta tesis. También, se comentan condiciones que se fuerzan a lo largo del juego.

2.4.1. Balanceo del tablero

Las propiedades de las cartas del tablero fueron elegidas y distribuidas de manera tal que haya cartas que se diferencien entre sí por no más de una o dos de ellas. Por lo tanto, hay cartas que para llegar a adivinarlas hay que pasar necesariamente por una mínima cantidad de preguntas. Así como también hay otras que son más simples de acertar, con dos o tres preguntas ya se puede llegar a la carta buscada.

En los casos en los que es necesario que el sujeto gane, se define que la carta a acertar sea fácil de adivinar. Caso contrario, se elige una carta que sea más complicada de alcanzar, es decir, que esa carta sea la última en quedar “levantada”. En las partidas finales, como no es relevante el resultado de la partida, se eligen las cartas al azar.

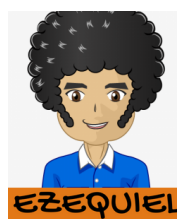


Figura 16: Ficha de la ronda 1 correspondiente al personaje Ezequiel



Figura 17: Fichas de la ronda 1 con pocas propiedades distintas a la carta del personaje Ezequiel

Personaje	Diferencia
Daniel	Sombrero / Color de ojos
Gustavo	Tipo de pelo
Pedro	Anteojos
Raúl	Tamaño de la nariz
Santiago	Largo del pelo

Tabla 2: Personajes de la ronda 1 y las propiedades válidas que lo diferencia de la carta Ezequiel

Veamos un ejemplo. Supongamos que el sujeto tenga que adivinar la carta de Ezequiel (Figura 16) en la primera ronda. Viendo las imágenes de un conjunto de cartas (Figura 17), se puede notar que Ezequiel se diferencia con cada uno de los otros personajes a partir de una o dos propiedades; considerando sólo las que son válidas en esta ronda. Se puede ver en la Tabla 2 cuáles son las propiedades que distinguen una carta de otra. De esta manera, si en el tablero del sujeto están disponibles una o más de estas cartas, además de la de Ezequiel, se requieren tantas preguntas como cantidad de cartas levantadas para llegar a la carta que se debe adivinar.

El diseño del tablero balanceado requirió un tiempo considerable debido a que se tenían que tener en cuenta muchos factores para que el juego quede lo más real posible y existan cartas que son “fáciles” y “difíciles” de adivinar.

2.4.2. Generación de preguntas

Se debió definir una estrategia para generar las preguntas que los avatares hacen en cada turno. A continuación describimos las alternativas.

2.4.2.1. Backtracking

A la hora de elegir cuál va a ser la siguiente pregunta que va a decir el avatar, en primera instancia se pensó en utilizar el algoritmo de **backtracking**. Se desarrolló el algoritmo para obtener todas las preguntas posibles a decir en el siguiente turno. A partir de ellas, se elegiría la mejor según convenga.

Con este algoritmo se intentó, para cada turno, ver a futuro todas las posibles preguntas y respuestas que se puedan dar entre el sujeto y el avatar en el resto del juego. Cada pregunta abre un conjunto nuevo de ramas posibles con una cantidad muy alta de opciones de pregunta, debido a la cantidad de cartas y propiedades disponibles. Esto lleva a que el algoritmo tenga una complejidad muy alta, y por lo tanto, se tarde mucho en seleccionar la siguiente pregunta. Por ello, no sirve para un juego que requiere hacer preguntas rápidamente, teniendo que mantener la dinámica.

Para poder reducir el tiempo de ejecución se aplicaron podas al algoritmo de backtracking. Sin embargo, no se notaron grandes diferencias. Otra propuesta fue reducir la profundidad del árbol del algoritmo. Sin embargo, para los primeros turnos del juego el conjunto de ramas generadas es muy alto, por lo cual disminuir su longitud ofrece una ventaja notoria pero no suficiente para el objetivo buscado.

También se consideró utilizar el algoritmo de backtracking del tercer turno en adelante para cada partida del juego. Aún así seguía sin resolverse en un tiempo óptimo. Se descartó entonces la utilización de este algoritmo.

2.4.2.2. Preguntas predefinidas y reglas según condiciones

Debido a la imposibilidad de aplicar el algoritmo de backtracking para generar las preguntas de los avatares, se establecieron ciertas reglas con respecto a cómo definir las.

Si necesitamos que el avatar gane una partida, se estableció que todas las preguntas que hagan estén predefinidas. Esto es debido a que podemos elegir una carta del tablero que requiera una mínima cantidad de preguntas para poder adivinarla. Esta carta es asignada al sujeto sabiendo que tarda al menos una cierta cantidad de preguntas, por lo tanto, el avatar debe hacer al menos esa cantidad. En este juego, en las partidas en las que el avatar deba ganar, la cantidad de preguntas predefinidas es seis.

Por otro lado, las primeras tres preguntas para el avatar son predefinidas si el sujeto debe ganar. El efecto de estas preguntas es bajar pocas cartas en el tablero del avatar, para así forzar al sujeto a que pueda ganar, teniendo que arriesgar la computadora en la mayoría de los casos. Las demás preguntas se eligen de manera aleatoria de un conjunto que cumpla ciertas reglas.

Una de las reglas es considerar las opciones que bajen una cantidad similar de cartas a la que bajó el sujeto. Para todas las posibles preguntas, se calcula cuántas cartas bajarían en el tablero. Teniendo en cuenta el número de cartas que bajó el sujeto, se obtienen las que tengan menor diferencia respecto a ese valor. También, si el avatar tiene menos cartas que el sujeto entonces se debe bajar la menor cantidad de cartas posibles. Aparte, de acuerdo a la pregunta anterior que realizó la computadora, se agrega al conjunto de opciones los valores que queden por preguntar de esa propiedad, en el caso que sea posible. Por último, no se pueden repetir preguntas ya hechas en turnos anteriores de esa misma partida.

En el caso de las partidas finales, donde no es necesario restringir quién debe ganar o perder, no se estableció ninguna pregunta predefinida. Todas las preguntas se eligen a partir de las reglas descriptas anteriormente.

2.4.3. Manipulación

2.4.3.1. Cantidad de preguntas necesarias

Se trataron de armar las partidas de manera tal que el sujeto pase por al menos seis turnos antes de que finalice. Éste es un buen número para contar con el material suficiente para el desarrollo del experimento y se logre producir el efecto deseado.

Si se requiriera mayor cantidad de preguntas y respuestas, el tablero tendría que haber tenido mayor cantidad de cartas, lo cual implica un espacio mayor en la interfaz. Como consecuencia, el juego también sería más extenso y tedioso.

Si se disminuye la cantidad de preguntas necesarias, existe el riesgo de que no se produzca la influencia del objetivo en el sujeto. Por lo que no es seguro que el sujeto sea influenciado por el efecto de la política de mimetización.

2.4.3.2. Resultados de las partidas

Como se contó previamente, es necesario que el sujeto gane y pierda la misma cantidad de veces contra todos los avatares. Por esa razón, se implementó un algoritmo para manipular el juego para poder cumplir con la condición.

Cuando el sujeto debe arriesgar un personaje, según si debe ganar o perder la partida, se resolvió qué hacer con la carta que deba adivinar. Se describen a continuación cada caso y cómo se resuelve.

- Si el sujeto tiene que ganar y al arriesgar acierta, se da vuelta la misma mostrando el personaje correcto.
- Si tiene que ganar y cuando arriesga no adivina la carta correcta, se da vuelta una carta diferente a la que tiene la computadora. La carta que se muestra es la que elige el sujeto al arriesgarse.

- Si tiene que perder y acierta la carta, se muestra una carta diferente a la correcta. Ésta es cualquiera de las que quedan en el tablero del sujeto.
- Si tiene que perder y no acierta la carta, se muestra la carta que tiene *en su mano* el avatar de la computadora.
- Si la computadora tiene que ganar, se define para que se arriesgue por la carta que efectivamente tiene el sujeto.
- Si tiene que perder, entonces elige cualquiera de las cartas que tiene disponibles en su tablero, descartando la que tiene el sujeto.

Este algoritmo nunca falló durante el experimento, es decir, que el sujeto siempre ganó o perdió según lo planeado.

2.4.4. Recuperación ante fallas

Pensando en situaciones en que se corte Internet, o cualquier otra situación que pueda llevar a que se rompa el juego, se decidió agregar la funcionalidad de poder comenzar un juego ya empezado. El sujeto debe volver a jugar la partida en la que el juego falló.

2.5. Mago de Oz

El Mago de Oz consiste en tener una persona que interprete en el momento lo que el sujeto va diciendo. La existencia del mismo permite al sujeto hablar libremente sin imponer restricciones en la gramática. También controla que el sujeto respete las reglas del juego. El Wizard of Oz se encuentra en otra computadora con otra interfaz propia monitoreando las acciones del sujeto. Los estudios Ter Maat et al. [2010] y Skantze and Hjalmarsson [2010] son casos de aplicación exitosa del Wizard of Oz.

En ninguno de los actos de diálogo del sujeto en el juego se requiere una interpretación exacta de lo que dice. Para cada uno es suficiente con saber lo imprescindible para continuar con el juego. Al describir los distintos botones de la interfaz del Wizard of Oz se explica en detalle cada caso.

Contando con la ventaja de que esta tesis la desarrollaron dos personas, se pudo llevar a cabo esta estrategia. Tener implementado un Mago de Oz implica que una persona debe estar frente a una computadora durante todo el experimento. Mientras tanto, la otra persona es la encargada de guiar al sujeto en las distintas etapas.

2.5.1. Interfaz



Figura 18: Interfaz del Wizard durante la ronda de práctica y la primera ronda del juego

El Wizard cuenta con una interfaz basada en botones para agilizar el proceso de traducción. La misma se puede ver en la Figura 18 y se divide en las siguientes secciones:

- **Los botones de las propiedades que el sujeto puede preguntar en esa ronda.** Si el sujeto pregunta “¿Tiene color de pelo negro el personaje de la carta que debo adivinar?” entonces el Wizard utiliza el botón de “pelo negro”. De esta manera, se trata de optimizar lo más posible el tiempo de respuesta del Mago de Oz para que se pueda mantener una dinámica de juego similar a la real. Luego, el avatar responde a la pregunta que el sujeto le dijo actuando acorde a lo que el Wizard traduzca.
- **Los botones especiales.** Luego de apretar uno de estos botones se muestra en la pantalla del sujeto, al lado del botón de audio, un cartel indicando al sujeto que realice una determinada acción. Esta sección abarca los siguientes botones:
 - **Repetir pregunta, Repetir respuesta, Repetir oponente, Repetir carta arriesgada.** Los botones de repetición se utilizan cuando no se comprende lo que el sujeto dijo. Ya sea porque no modula bien, habla bajo o el audio se corta. Otro motivo es cuando el sujeto no formula la oración con el largo indicado.
 - **No podés arriesgarte / Debes arriesgarte.** Según el momento de la partida se muestra alguno de los dos botones. En casi todos los turnos se muestra el primero, salvo cuando al avatar le queda una única carta y el sujeto en su turno debe arriesgarse por una carta obligatoriamente. El primer botón, se utiliza cuando el sujeto intente arriesgarse cuando aún le quedan cartas por bajar del tablero. El otro se utiliza cuando en el tablero del sujeto solo quede una carta y este no se arriesgó en su turno.
 - **Pregunta inválida.** Este botón es utilizado cuando el sujeto pregunta por alguna propiedad de la carta que no está entre las permitidas de la ronda actual. El sujeto tiene que formular una nueva pregunta usando alguna de las propiedades posibles.
- **El botón de “Respondió”.** No es necesario para el sistema saber la respuesta del sujeto ya que se le pide que no mienta, por ende la respuesta ya es conocida. Pero sí se requiere saber cuándo termina de responder. Como consecuencia, se bajan las cartas del tablero del avatar y se pasa al turno del sujeto.
- **El botón de “Respondió al riesgo”.** Este botón se utiliza cuando el sujeto responde a la pregunta en la que el avatar se arriesga por una carta. Al igual que con las respuestas, no nos interesa lo que diga ya que no influye en el resultado de la partida. Este botón implica que se termina la partida, por lo que se le muestra al sujeto el cartel con el resultado de la misma y la tabla de posiciones hasta ese momento en esa ronda.

- **Lista de los nombres de las cartas.** Cuando en el tablero del avatar queda una única carta, el sujeto está obligado arriesgarse por una. El Wizard escucha el nombre de la carta y hace clic sobre ese nombre en el listado. Éste muestra únicamente los nombres de las cartas que están “levantadas” en el tablero del sujeto. Como efecto de seleccionar una carta la partida finaliza.
- **Lista de los nombres de los avatares del juego.** Cuando el sujeto termina de jugar las cuatro partidas de la ronda, debe elegir contra quién quiere jugar la partida final. El Mago de Oz selecciona el nombre del avatar elegido. En la pantalla del sujeto se destaca al mismo.
- **Carteles de acciones del sujeto.** Para las acciones más importantes que realiza el sujeto se muestran carteles informativos en la pantalla del Wizard. Esto es necesario para que el Mago de Oz pueda seguir el juego y controlar correctamente lo que va haciendo el sujeto. El Wizard no puede guiarse únicamente por los audios pues puede ocurrir que el sujeto pregunte cuando debería estar respondiendo. Por eso, se decidió agregar los carteles orientativos. Por ejemplo:
 - “*El sujeto terminó de grabar*”: Este cartel se muestra cuando el sujeto suelta el botón de audio.
 - “*El sujeto debe RESPONDER*”: Este cartel se muestra cuando el sujeto tiene en el botón de audio la etiqueta de “Responder”.

En la primer prueba piloto, el Wizard no tenía manera de asegurarse de haber tocado los botones. Por ese motivo, se decidió agregar una funcionalidad en la que los botones titilen cuando se toca alguno.

2.5.2. Comunicación con el juego

Mientras el sujeto mantiene apretado el botón de audio se envían los datos correspondientes al audio al host del Wizard. Se resolvió utilizar el protocolo Transmission Control Protocol (TCP) para el envío de la información, ya que es confiable en la transmisión de paquetes. Por otro lado, es necesario contar con una red con una IP fija para cada host, que permita la conexión entre la computadora del sujeto y la del Mago de Oz.

La comunicación es bilateral puesto que desde la computadora del sujeto se envía el audio y cierta información relevante para que el Wizard pueda desarrollar su tarea. Por otro lado, el Mago de Oz envía al host de la computadora con el juego las indicaciones necesarias acordes a lo escuchado en los audios. Se ilustra en la Figura 19.

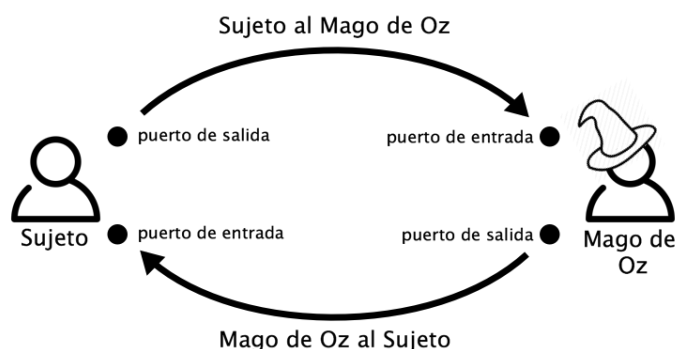


Figura 19: Comunicación bilateral entre los hosts del Mago de Oz y del sujeto

2.6. Procesamiento de los atributos acústicos-prosódicos

En esta sección describimos las tareas de procesamiento de los audios grabados por los sujetos y los sistemas que sintetizan las voces de los avatares.

2.6.1. Procesamiento del audio

2.6.1.1. Reconocimiento del habla

Los atributos de la prosodia que se tienen en cuenta en el experimento son el *tono*, la *tasa del habla* y la *intensidad*. A continuación, se describen los distintos sistemas y procedimientos para obtenerlos.

Se investigaron distintos servicios que brindan la funcionalidad de *speech-to-text* y se eligió integrar la aplicación de IBM⁵. Una vez recibido el audio de la pregunta o respuesta que haya hecho la persona, se envía el mismo al reconocedor de IBM. La transcripción que devuelve este sistema se utiliza en un script de Praat⁶ para poder calcular los atributos prosódicos del audio.

Praat es un software utilizado para el análisis del habla. Permite el análisis de la entonación, la intensidad, entre otros atributos de la prosodia. Praat se basa en los valores del tono en los que se estima que está el habla humana. Estos valores están en un rango standard de 75 a 500 Hz. Para mujeres se recomienda un rango entre 100 - 500 Hz. En cambio, para los hombres el rango será de 75 - 300 Hz. Por este motivo, es necesario diferenciar en el experimento el género del sujeto.

Para cada **segmento** (que luego explicaremos en más detalle) se aplica un script de Praat y se obtienen el tono y la intensidad. Con estos valores se calcula el promedio ponderado, el cual es utilizado para las futuras sintetizaciones.

La tasa del habla se calcula a partir de la cantidad de sílabas y la duración del audio. Para calcularla no es necesario un reconocedor que transcriba exacto lo dicho por la persona, alcanza con respetar la cantidad de sílabas. El sistema de reconocimiento del habla utilizado en este trabajo, IBM, devuelve no sólo la transcripción reconocida, sino también los timestamps de cuándo se dijo cada una de las palabras reconocidas dentro del audio. En base a estas transcripciones, se puede obtener el número de sílabas de cada palabra. Para esto se utiliza el *syllabification algorithm* [Hernández-Figueroa et al., 2013]. Llamamos **segmento** a un conjunto de palabras sin silencios entre ellas. A partir de lo obtenido del algoritmo de silabificación, para cada segmento se calcula la tasa del habla como la relación entre la cantidad de sílabas dichas en ese segmento de habla y su duración, la cual es medida en segundos. En caso de haber más de un segmento hablado dentro de una grabación, se toma como valor final de la tasa del habla al promedio de las distintas tasas de habla calculadas, ponderado por su duración.

Para cada segmento también se calcula el tono y la intensidad y luego, el promedio ponderado. La Fórmula 1 es utilizada para obtener el promedio ponderado del tono y de la intensidad, mientras que la Fórmula 2 es para obtener la tasa del habla.

$$prom_ponderado(atributo) = \frac{\sum_{seg} longitud_{seg} * atributo_{seg}}{\sum_{seg} longitud_{seg}} \quad (1)$$

$$tasa_del_habla = \frac{\#sílabas}{\sum_{seg} longitud_{seg}} \quad (2)$$

⁵<https://www.ibm.com/watson/services/speech-to-text/>

⁶<http://www.fon.hum.uva.nl/praat/>

Se puede observar el siguiente ejemplo en el que un sujeto pregunta: “¿El personaje de tu carta tiene anteojos?”. Del reconocedor se obtienen los siguientes valores:

```
[ [u'el', 0.64, 0.78], [u'personaje', 0.78, 1.4], [u'de', 1.4, 1.55],
[u'tu', 1.55, 2.29], [u'carta', 2.32, 2.73], [u'tiene', 2.73, 3.05],
[u'anteojos', 3.05, 3.73] ]
```

Cada uno de los elementos está formado por una palabra, los tiempos de inicio y de fin de dicha palabra en el audio.

Observando los timestamps finales de las palabras se pueden observar que hay uno en el que no coincide con el inicial de la siguiente palabra. Esto quiere decir que ahí hubo un silencio y que debemos ignorarlo para el cálculo de los atributos. El único silencio que se nota en el ejemplo es el que está entre las palabras “*tu*” y “*carta*”. Por lo tanto, los segmentos que se forman son:

```
[[u'el personaje de tu', 0.64, 2.29], [u'carta tiene anteojos', 2.32, 3.73]]
```

Luego, se aplica el script de Praat para cada segmento, se obtienen los valores de la Tabla 3.

Segmento	Intensidad (dB)	Tono (Hz)	Cantidad de sílabas	Inicio (seg)	Fin (seg)	Duración total (seg)
el personaje de tu	81,12	179,56	7	0,64	2,29	1,65
carta tiene anteojos	80,95	177,34	8	2,32	3,73	1,41

Tabla 3: Valores de los segmentos

Por lo tanto, la tasa del habla en este ejemplo es:

$$\frac{7 \text{ sílabas} + 8 \text{ sílabas}}{1,65 \text{ seg} + 1,41 \text{ seg}} = 4,9 \text{ sílabas/seg} \quad (3)$$

La intensidad es:

$$\frac{81,13 \text{ dB} * 1,65 \text{ seg} + 80,95 \text{ dB} * 1,41 \text{ seg}}{1,65 \text{ seg} + 1,41 \text{ seg}} = 81,045 \text{ dB} \quad (4)$$

El tono es:

$$\frac{179,56 \text{ Hz} * 1,65 \text{ seg} + 177,34 \text{ Hz} * 1,41 \text{ seg}}{1,65 \text{ seg} + 1,41 \text{ seg}} = 178,537 \text{ Hz} \quad (5)$$

2.6.1.2. Síntesis del habla

En un principio, se comenzó el desarrollo utilizando un sintetizador basado en Mary. El framework open-source MaryTTS⁷ cuenta con una voz en español argentino, desarrollada como parte de la Tesis de Licenciatura de Luisina Violante [Violante et al., 2013]. Usa síntesis basada en HMM, entrenado con datos de una locutora profesional en español rioplatense. Sin embargo, el sintetizador no funciona correctamente para las preguntas, es decir, al escuchar el audio no se aprecia el tono de pregunta. Por otro lado, la voz del sintetizador en Mary no suena natural.

Luego, se utilizó como servicio de text-to-speech el que ofrece IBM⁸. Este sistema presenta la desventaja de que las preguntas no se escuchan con tono interrogativo. Por esto, se tomó la decisión de que las preguntas del juego se debían hacer con formato de consulta, es decir preguntas indirectas. Se forzaba al sujeto a cumplir esta condición para que la comunicación sea similar desde ambos lados.

⁷<http://mary.dfki.de>

⁸<https://www.ibm.com/watson/services/text-to-speech/>

En consecuencia, optamos por el sistema TTS Polly, ofrecido a través de la plataforma Amazon Web Services⁹, en el cual las preguntas sintetizadas se escuchan de la manera deseada. Así, no se tiene que complicar al sujeto forzando a que no pueda hacer preguntas directas. Por otra parte, no se pierde la espontaneidad del juego de ir haciendo preguntas para averiguar al personaje que tiene el avatar para ganar.

El servicio de Amazon ofrece sintetizar un determinado texto, pudiendo manipular los valores de tasa del habla y tono. El sistema permite manipular estos atributos en base al porcentaje de variación respecto al nivel base del sistema, lo cual es positivo para los fines de este trabajo. Por ejemplo, se puede solicitar que sintetice el texto un +10 % más agudo o rápido que lo que lo hace por default. La intensidad se modifica utilizando *SoX*¹⁰.

Además, este TTS retorna los audios con valores de atributos prosódicos bastante próximos a los indicados. Se realizó un script que sintetiza cuatro frases distintas con múltiples combinaciones de los tres atributos prosódicos que se analizan en esta tesis. Luego, se utilizó el mismo reconocedor del juego para observar cuáles son los valores de los atributos medidos. La relación entre la variación deseada y la que se obtiene sintetizando con Polly debería ser lineal y con pendiente uno. Efectivamente, se observa ese comportamiento en los gráficos de la Figura 20.

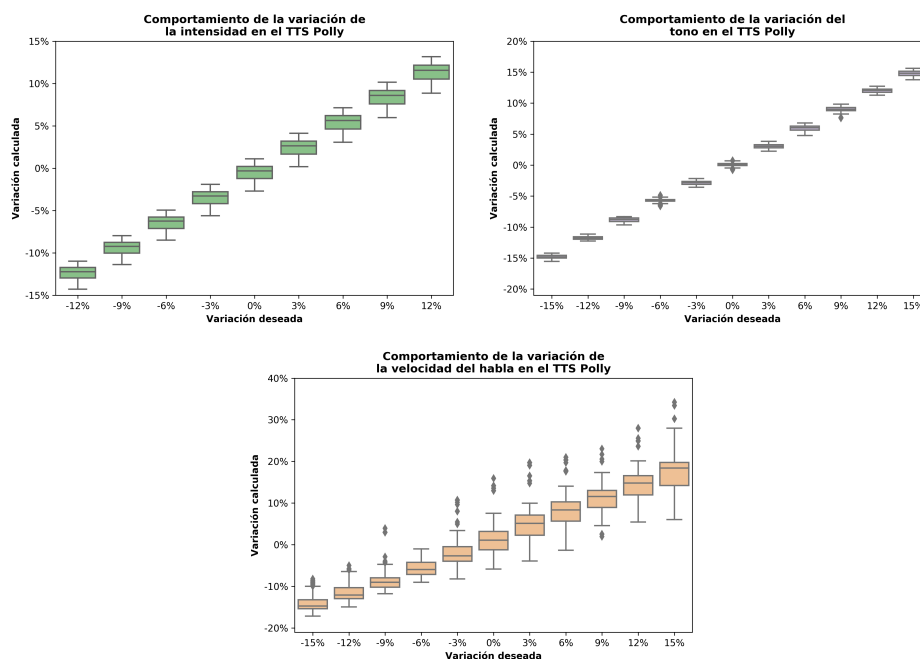


Figura 20: Comparación entre la variación deseada y la obtenida con el sintetizador de Amazon

2.6.1.3. Características del habla del sujeto

Antes de comenzar el experimento, se le da al sujeto una serie de indicaciones con respecto a su habla. Primero, se le pide que hable de manera natural para poder obtener atributos del habla reales y no sobreactuados.

Asimismo, se le indica que formule oraciones lo más completas posibles tanto en preguntas como respuestas. En el juego de mesa, las preguntas suelen ser cortas y puntuales, y las respuestas, monosílabos. En las instrucciones del juego se le muestra al sujeto ejemplos de preguntas y respuestas para que las considere al momento de hablar. Esta necesidad se debe al cálculo del speech rate. Para lograr el mismo, se utiliza el tiempo que tarda en decir la oración el sujeto, ignorando los silencios, y la cantidad de sílabas que dijo. Las oraciones más extensas permiten lograr un cálculo más preciso de

⁹<https://aws.amazon.com/polly/>

¹⁰<http://sox.sourceforge.net/>

este valor.

2.6.2. Speech history

En los cálculos para obtener los valores de los atributos con los que habla el avatar son necesarios valores calculados previamente. Para esto, se mantiene una estructura denominada **speech history**. Ésta almacena todos los atributos del habla de los audios del sujeto, tanto preguntas como respuestas. Como ya se mencionó, los atributos analizados en este experimento son la tasa del habla, tono e intensidad. Éstos se obtienen luego de aplicar la serie de cálculos y scripts de Praat a los valores devueltos por el reconocedor de IBM (descriptos en la Sección 2.6.1.1).

La estructura del speech history es guardada en un archivo luego de finalizar cada partida. De este archivo se obtienen los valores de la partida anterior para los cálculos de los atributos prosódicos en los casos que es necesario volver a comenzar el juego en alguna partida del medio. Estos casos se pueden dar luego de que haya un error en el juego (ver Sección 2.4.4).

2.6.3. Actos de diálogo

Como se presentó en la introducción, en cada pregunta y respuesta de la conversación del juego se utiliza una política de mimetización, según corresponda. En la ronda de práctica se utiliza el habla neutra. Para lograr este efecto de mimetización, son necesarios algunos cálculos antes de utilizar el sistema de TTS para el avatar correspondiente.

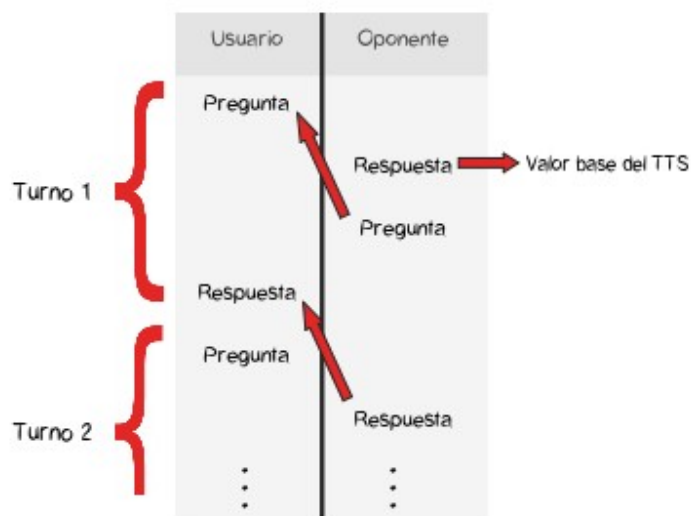


Figura 21: Diagrama de obtención del valor anterior del atributo prosódico para cada acto de diálogo

Previo a la síntesis se debe calcular los valores de los tres atributos prosódicos que se analizan en este proyecto. Cada uno se obtiene a partir del valor base de esa partida y el valor extraído del audio del sujeto en la acción (pregunta/respuesta) anterior correspondiente.

El *valor base* de la fórmula para cada atributo se obtiene del promedio de los valores de la partida anterior para esa misma acción. Estos valores son extraídos del *speech history*.

La diferencia de si el resultado aplica política de entrainment o disenchantment lo va a marcar el valor *changer* de la fórmula 7. Si *changer* es 0 entonces el efecto es el del habla neutra, ya que no se aumenta ni disminuye el valor default de los atributos prosódicos. Si *changer* es -1 entonces se aplica disenchantment en la voz del avatar, mientras que si es 1 entonces se aplica entrainment.

$$porcentaje = \frac{valor\ del\ audio - valor\ base}{valor\ base} * 100 \quad (6)$$

$$valor\ parcial = porcentaje * changer \quad (7)$$

Se establecen valores límites mínimo y máximo para cada atributo prosódico, y de esta manera, evitar que la voz del avatar suene poco natural. Si el resultado obtenido en la fórmula 7 no está dentro del rango que definen los límites, se toma el límite superior o inferior según corresponda a la fórmula 8.

$$resultado = \max(\min(valor\ parcial, \textit{límite máximo}), \textit{límite mínimo}) \quad (8)$$

Atributo	Límite máximo (%)	Límite mínimo (%)
Tono	8	-8
Intensidad	15	-15
Tasa del habla	30	-30

Tabla 4: Límites máximos y mínimos definidos en el experimento para los atributos prosódicos

Luego de estos cálculos se obtienen los porcentajes de aumento o disminución del tono, la tasa del habla e intensidad. Los mismos se aplican a la voz por defecto que usamos del TTS de Amazon.

3. Resultados

En la Sección 3.1 explicamos el cómo y dónde se llevaron a cabo los experimentos; además, se mencionan todos los recursos necesarios. Luego, se analizan los resultados obtenidos. En la Sección 3.2 se estudian según la política de mimetización de las elecciones de los sujetos. En la Sección 3.3 se hacen diversas agrupaciones para profundizar. A continuación, en la Sección 3.4 se analizan las elecciones de los sujetos según el resultado de la última partida de la ronda. Finalmente, en la Sección 3.5 se hace un análisis de potencia teniendo en cuenta los resultados de la Sección 3.2.

3.1. Recolección de datos

Los experimentos del juego se llevaron a cabo durante el mes de Agosto del 2017 en el Laboratorio de Graduados del Departamento de Computación de la Facultad de Ciencias Exactas y Naturales. Se lo eligió porque es el más silencioso de los laboratorios del departamento, sin ventanas al exterior. La interfaz del sujeto siempre se levantó en la misma computadora, una de las del laboratorio nombrado. La interfaz del Mago de Oz se corrió en una notebook conectada a la misma red. La notebook cuenta con pantalla táctil para poder agilizar el tiempo de respuesta luego de escuchar al sujeto. El Mago de Oz estuvo situado en el mismo laboratorio pero el sujeto se encontraba de espaldas durante el experimento, para evitar que pudiera ver la pantalla de la notebook.

Para el experimento se requirieron dos personas, una encargada de ser el Mago de Oz y la otra responsable de guiar a los sujetos. Estas personas no cambiaron sus roles en todos los muestreos. Para poder jugar los sujetos contaron con un headset, un mouse y una cheat sheet (ver Apéndice C.1) con las explicaciones de las etiquetas del botón de audio y estructuras de ejemplo de preguntas y respuestas. El Wizard of Oz usó el mismo modelo de headset que los sujetos.

El experimento lo hicieron 16 personas diferentes (8 mujeres, 8 hombres), de entre 18 y 65 años, con un promedio de edad de 29,75 años y un desvío estándar de 12,075 años. En la Figura 22 se observan las edades y sexo de los participantes del experimento *¿Quién es quién?*.

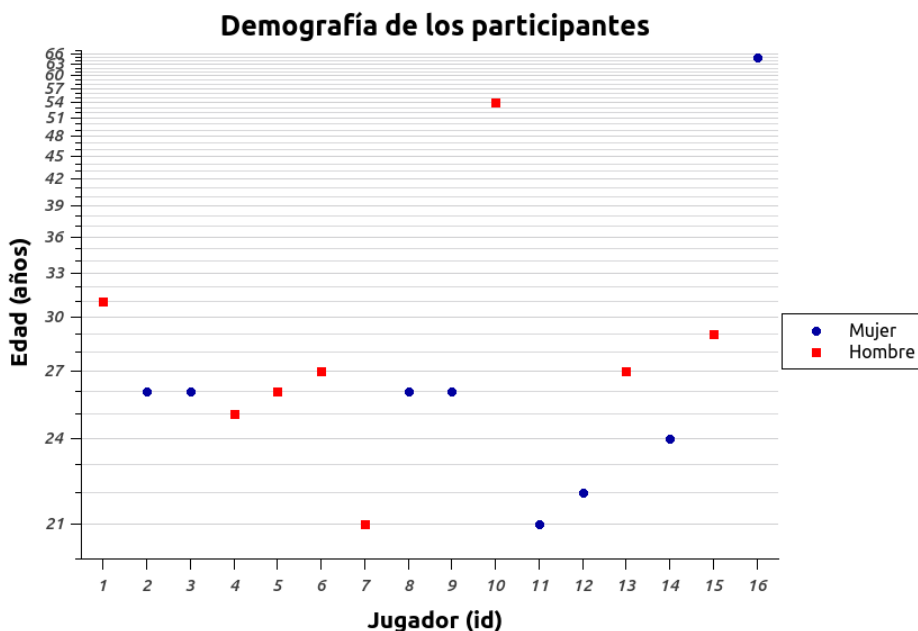


Figura 22: Edades de los participantes del experimento separados por género

Por cada modo de juego se muestrearon dos hombres y dos mujeres, a fin de mantener balanceada la muestra. Debido a la larga duración del experimento y la cantidad de personas requerida para llevarlo a cabo, no fue posible tomar una mayor cantidad de muestras.

Los sujetos fueron remunerados por su tiempo, recibiendo \$90 por hora. El tiempo promedio de juego fue de 4780,25 segundos (una hora y veinte minutos aproximadamente). Además, se dio un bonus extra por cada partida final ganada para fomentar la importancia de la elección del avatar. Antes del comienzo del juego, se les recalcó a los sujetos el beneficio de prestar atención a qué avatar sería el contrincante menos inteligente o hábil. Elegir un avatar al que consideren capaces de ganarle les generaría más ganancias. Este incentivo fue necesario para intentar tener la atención del sujeto a la hora de la elección.

Esta investigación fue aprobada por el Comité de Ética del CEMIC (Centro de Educación Médica e Investigaciones Clínicas Norberto Quirno), protocolo titulado “Relationship between trust and entrainment in speech”, Investigador Principal: Dr. Agustín Gravano, aprobado el 20/11/2014 y renovado anualmente el 4/12/2015 y el 21/11/2016.

Antes de que los sujetos comiencen a jugar se les leyó un texto introductorio, disponible en el Apéndice C.2. Se les dio un consentimiento de conformidad (ver Apéndice C.3) que debieron firmar luego de leerlo. Se les recordó que el experimento es anónimo y confidencial y que podían abandonarlo cuando lo desearan. Acto seguido, llenaron un formulario de personalidad (ver Apéndice C.4). Posteriormente, el sujeto leyó las instrucciones del juego (ver Apéndice C.5) y pasó a sentarse frente a la computadora a jugar.

Una vez finalizado el juego, los sujetos respondieron un segundo formulario con preguntas respecto a su formación, disponible en el Apéndice C.6. Finalmente recibieron la remuneración y firmaron el recibo (ver Apéndice C.7) que indica el valor del dinero entregado. Antes de irse se les entregó una hoja con detalles y el contexto del experimento, que se puede leer en el Apéndice C.8.

Vale aclarar que, antes de llevar a cabo el experimento, se realizaron tres etapas de pruebas piloto para testear el correcto funcionamiento del juego y su dinámica. Gracias a las mismas pudieron corregirse errores y mejorar la interfaz y otros detalles.

Luego de la recolección de datos, los logs¹¹ de cada juego fueron utilizados para el análisis del comportamiento de los sujetos en el experimento.

¹¹<https://docs.python.org/2/library/logging.html>

3.2. Análisis según la política de mimetización

Para empezar, nos pareció interesante estudiar cuántas veces se optó por un avatar que sigue la política de entrainment contra las que elecciones de uno con disentrainment. Se unifican todas las elecciones independientemente de las rondas, los modos de juego y los sujetos. De esto surge la Figura 23.

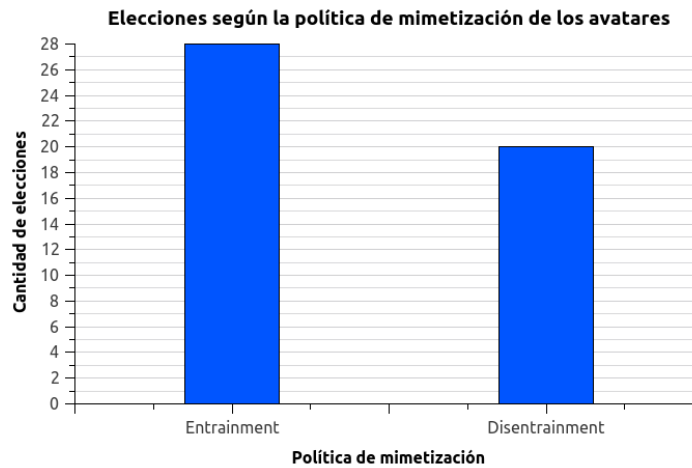


Figura 23: Comparación de la cantidad de veces que se eligió *entrainment* contra *disentrainment*

Es esperable que los sujetos elijan al avatar que consideren menos competente. Si una política de mimetización es elegida más veces, se puede ver un efecto positivo de la otra política. Se observa que se eligió una mayor cantidad de veces a un avatar con habla entrainment para jugar la partida final. Entonces, conjeturamos que esto se debe a que consideren más competente a los que hablan con disentrainment.

Los datos también fueron analizados a través de un test binomial. Consideramos nuestra hipótesis nula a “la media de la proporción de veces que eligió *entrainment* es 0,5”. Establecemos el nivel de significación en 0,05. Supongamos que la hipótesis nula es cierta, ¿cuál es la probabilidad de que una distribución binomial de parámetro 0,5 repetida 48 veces nos dé un resultado como el obtenido (28 entrainment y 20 disentrainment)? Calculando esta probabilidad llegamos a que el p -valor para este caso es igual a 0,3123. Este resultado no es inconsistente con la hipótesis nula, por lo que no se puede descartar que las elecciones hayan sido producto del azar.

3.3. Análisis de subgrupos

Por otro lado, nos pareció interesante analizar los datos agrupándolos según diferentes condiciones. Una manera es separar la totalidad de las elecciones por el género de los sujetos, como podemos observar en la Figura 24. Los hombres y las mujeres podrían comportarse de manera diferente frente a avatares de género femenino, como es en el caso de este proyecto. Se puede notar que en el caso de las mujeres no hubo una inclinación hacia ninguna política de mimetización. En cambio, los hombres tuvieron una clara preferencia por entrainment.

Se ejecutó un test binomial para cada agrupación de los datos, en el caso de las mujeres se obtuvo un p -valor igual a 1, por lo que no se puede descartar el azar. Por otro lado, en la agrupación por hombres se obtiene un p -valor igual a 0,15. Sin embargo, es más difícil obtener significancia estadística porque cada corte implica menos observaciones. En este caso, sólo se tiene 24 muestras por agrupación.

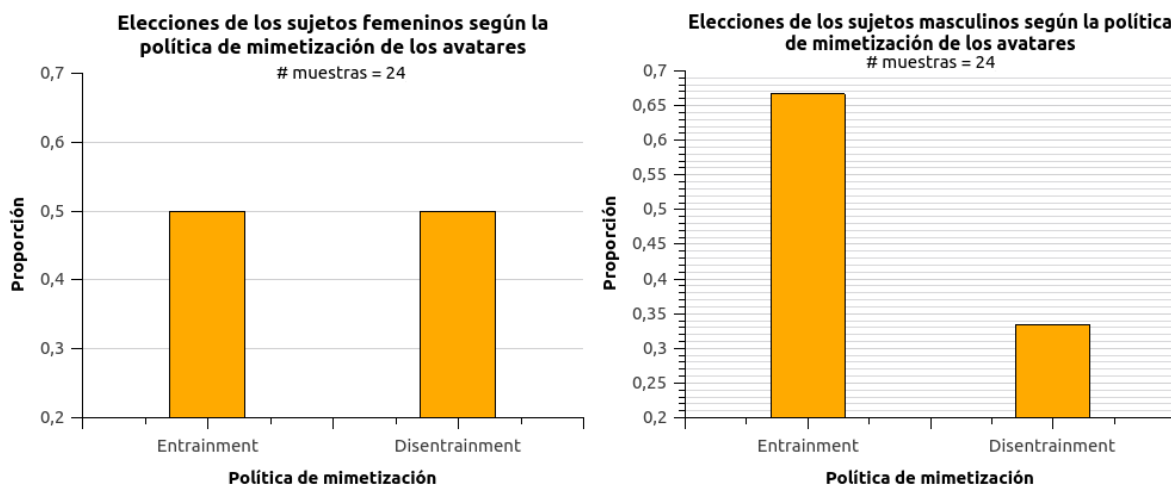


Figura 24: Proporción de las mujeres y de los hombres que eligieron *entrainment* contra *disentrainment*

Por otra parte, también asociamos los resultados de las elecciones por ronda del juego. Esto se debe a que en cada ronda del juego hay dos avatares diferentes cuyos nombres e imágenes podrían afectar a las elecciones de los sujetos. Se puede observar que la mayoría de los sujetos eligió *entrainment* en la primer y segunda ronda, mientras que en la última ronda no hubo preferencias con respecto a la política de mimetización. El p -valor obtenido para la primer y segunda ronda es 0,46, y para la tercera ronda es 1. Como en el caso anterior, no se puede descartar el azar y es complejo obtener una significancia estadística con 16 muestras.

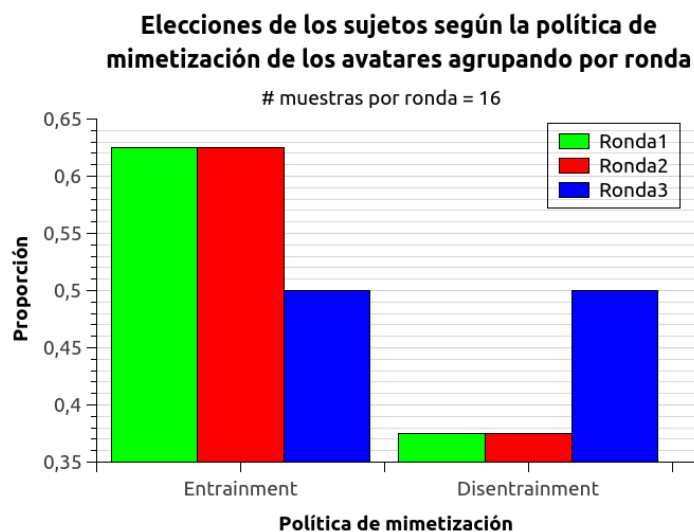


Figura 25: Proporción de los sujetos que eligieron *entrainment* contra *disentrainment* por ronda

Para refinar la agrupación anterior se separó por género de los sujetos. En los hombres hubo una preferencia por *entrainment*, pero se puede ver más marcada la diferencia en la segunda ronda. En el caso de las mujeres se observa que la mayoría optó por jugar contra un avatar con habla *entrainment* en la primer ronda, mientras que en la tercera ronda ocurrió lo contrario. Por otro lado, en la segunda ronda no se destaca ninguna política de mimetización. De acá en adelante, no se presentan los p -valores debido a que no se cuenta con suficiente cantidad de observaciones para poder justificar ese tipo de análisis.

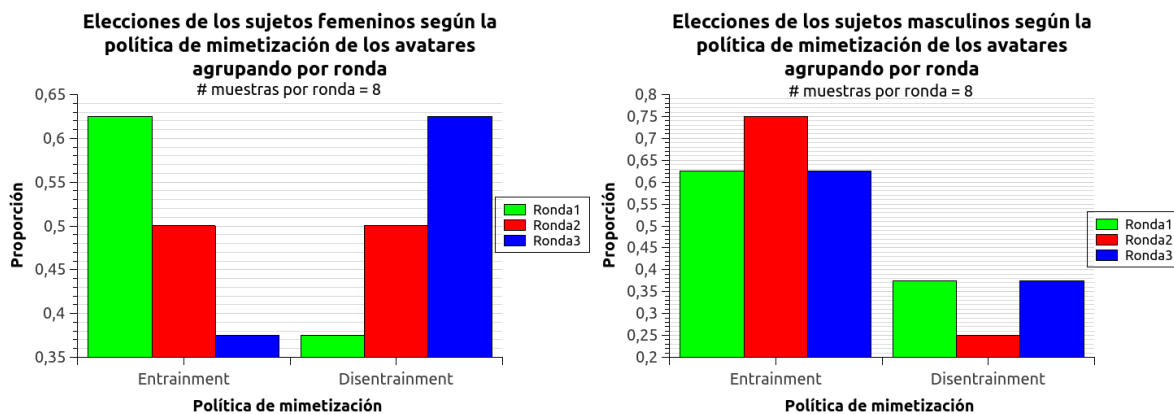


Figura 26: Proporción de los sujetos que eligieron *entrainment* contra *disentrainment* por ronda y género

Luego, como explicamos en la Sección 2.2.3, existen cuatro modos de juego que se diferencian principalmente por la política de mimetización y la imagen de los avatares. Se analizan las elecciones de los sujetos separándolas por modo. El objetivo de este análisis es ver si las variantes de cada uno efectivamente influyen en las decisiones de los sujetos. La preferencia más destacada es la del modo de juego 4. En el mismo se puede ver a *entrainment* como la política de mimetización más elegida ya que el 91,7% eligió al avatar con este efecto.

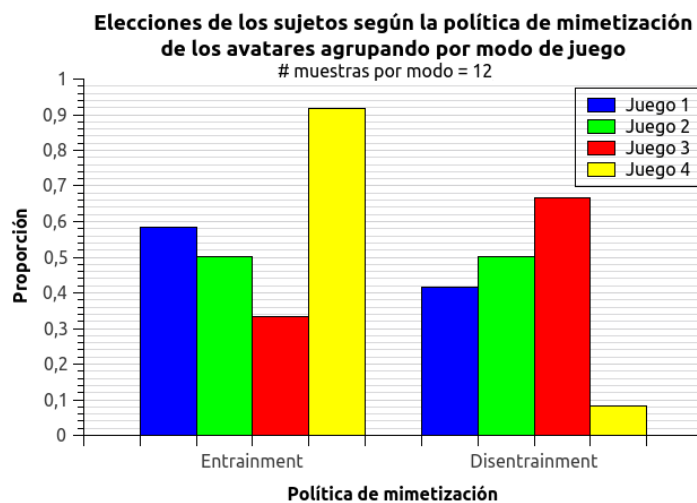


Figura 27: Proporción de los sujetos que eligieron *entrainment* contra *disentrainment* por modo de juego

Al igual que la agrupación por rondas, en este caso también se analizaron los datos distinguiendo por género. En el caso de los hombres se observó que en el modo de juego 4, el total de ellos eligió *entrainment*. Para más detalle, los gráficos se pueden ver en el Apéndice D.

Otro gráfico para analizar es la separación de las elecciones por jugador (ver Figura 28). Se quiere ver las preferencias de los sujetos individualmente por juego. Se destaca que cinco de las personas eligieron en su totalidad *entrainment* y sólo uno de los 16 eligió en las tres rondas *disentrainment*. Del resto se observa que la mayoría eligió una vez a un avatar con *entrainment* y dos veces al avatar con *disentrainment*. Sólo tres de los jugadores optaron más veces por *entrainment*. Por ende, la mitad de los usuarios prefirieron en su mayor parte *entrainment* mientras que la otra mitad optó por *disentrainment* en su mayoría.

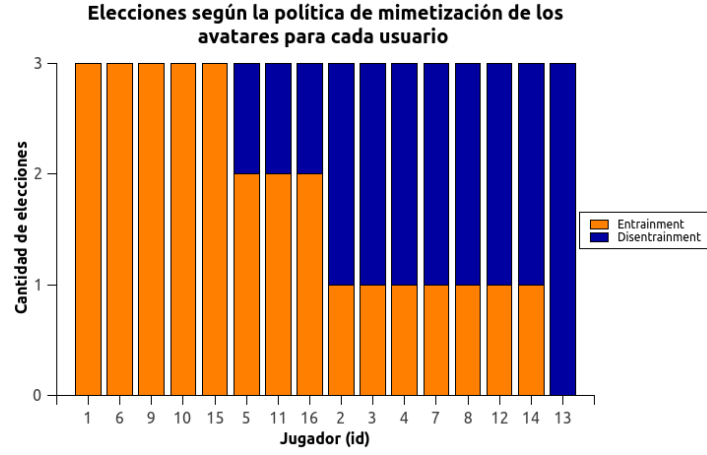


Figura 28: Proporción de las elecciones de mimetización por sujeto

3.4. Análisis según el resultado de las últimas partidas de cada ronda

En la Sección 2.3.6, discutimos que puede ocurrir que los sujetos hayan elegido al avatar contra el cual jugar la partida final según el resultado de la última partida de la ronda. Esta observación fue hecha por un sujeto de una de las pruebas piloto. Nos pareció interesante ver si esa condición se cumplió o no en los experimentos. Cumplir la condición implica que se haya tenido en cuenta el resultado de la última partida de la ronda en la elección del avatar: si en dicha partida el sujeto le gana al avatar, elige jugar contra éste, pero si pierde, elige jugar contra el otro avatar de esa ronda.

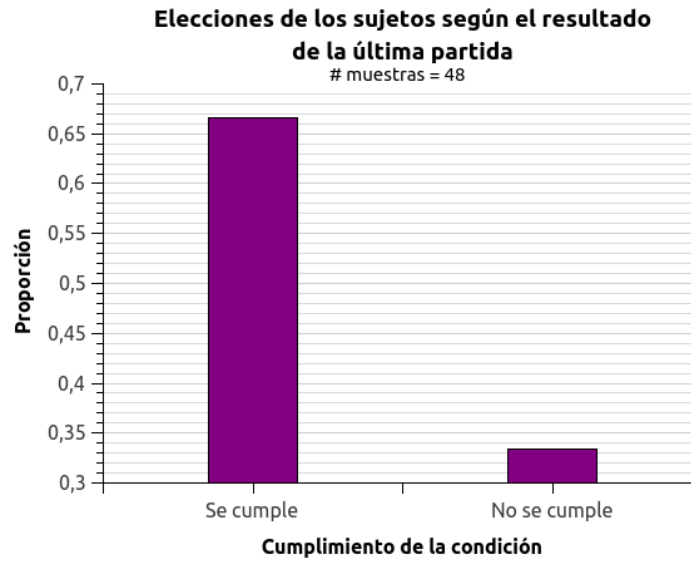


Figura 29: Proporción de los sujetos que cumplen la condición

La Figura 29 muestra que hay una proporción alta de sujetos que eligieron a un avatar contra el que habían ganado la última partida de la ronda, o al otro en caso de haber perdido.

Para este caso el p -valor arrojado por el test binomial con una probabilidad de 0,5 es de 0,0293. Esto quiere decir que estamos lejos de que los resultados sean producto del azar. Por ello, nos pareció pertinente hacer un análisis sobre esta condición separando por rondas.

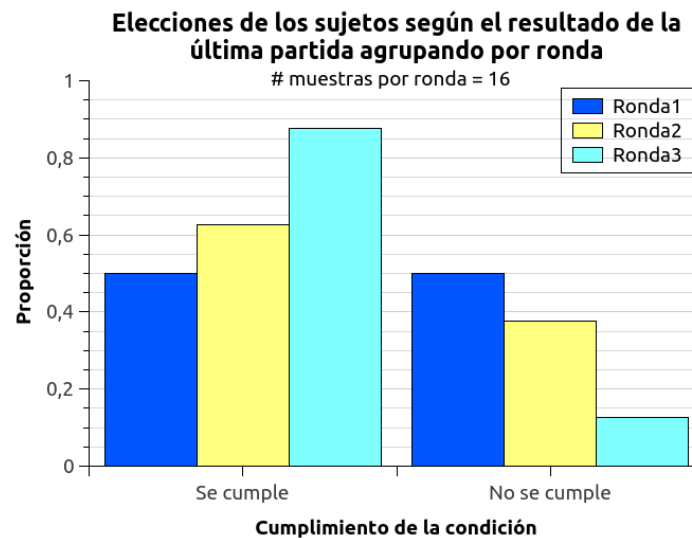


Figura 30: Proporción de los sujetos que cumplen la condición agrupando por ronda

Las imágenes de la Figura 6 de la Sección 2.2.4.2 resumen la organización de cada ronda del juego. En la primera y segunda ronda siempre ocurre que el sujeto gana la última partida sin importar el avatar. En cambio, en la tercera ronda el sujeto siempre pierde. Ésto es algo determinante en el análisis de esta condición.

En la Figura 30 se observa que en las primeras dos rondas no hay una marcada diferencia entre la elección del sujeto según el resultado de la partida. Caso contrario, en la tercera ronda hay más sujetos que eligieron al avatar de la última partida si le ganaron, o al otro si perdieron. Sin embargo, no quiere decir que su elección esté condicionada por cómo le fue en la última partida.

3.5. Análisis de potencia

Para poder descartar que las elecciones de los sujetos sobre una política de mimetización fueran azar o no, es necesario tomar un número mayor de muestras y así tener un número mayor de resultados. A modo ilustrativo, en la Tabla 5 veamos que ocurre si multiplicamos proporcionalmente la cantidad de sujetos y sus elecciones.

Potencia	Cantidad de elecciones	Cantidad de elecciones para entrainment	Cantidad de elecciones para disenitainment	<i>p</i> -valor
1	48	28	20	0,3123
2	96	56	60	0,1253
3	144	84	90	0,0549

Tabla 5: Test de potencia basado en las elecciones obtenidas en el experimento

A medida que se va aumentando proporcionalmente la cantidad de elecciones, se observa una disminución del *p*-valor, acercándose cada vez más al nivel de significación mencionado en la Sección 3.2. Por lo tanto, sería necesario aumentar la cantidad de muestras para poder llegar a una conclusión más certera.

4. Conclusiones y trabajo futuro

El objetivo principal de este proyecto es ver qué efecto produce sobre los sujetos la mimetización prosódica mediante sincronización. Se desea conocer qué política de mimetización asocian con una mayor nivel de competencia o habilidad, es decir, si los sujetos muestran preferencias por entrainment o disentrainment. Para ello, se desarrolló el juego “¿Quién es quién?”, para obtener un corpus de datos para analizar el habla y el comportamiento de las personas. El juego consta de tres rondas, en cada una de las cuales los sujetos juegan dos veces contra dos contrincantes. Estos últimos son avatares gemelos, uno con habla con entrainment y el otro con disentrainment. Terminada cada ronda, los sujetos debieron elegir contra cuál de los avatares jugar una partida final bonificada.

A partir de los resultados, se puede notar una tendencia a considerar como más *competente* a los avatares con habla con **disentrainment**. Esto surge de las elecciones de los sujetos, ya que la mayoría optó por jugar alguna o todas las partidas finales contra avatares de habla entrainment. Sin embargo, la tendencia descubierta es leve ya que sólo el 58,33 % eligió entrainment, lo que implica que no es significativa. Cabe destacar que en el análisis por jugador (ver Figura 28) la mitad eligió en su mayoría entrainment y la otra mitad optó más veces por disentrainment. Por lo tanto, no podemos asegurar que los sujetos consideren más capaces/competentes a los interlocutores que aplican la política de mimetización disentrainment.

En este trabajo sólo participaron del juego 16 personas, de las cuales la mitad eran hombres y la otra mujeres. Por ende, en total se obtuvieron 48 elecciones. Consideramos que este número no es suficientemente alto como para sacar conclusiones precisas. No fue posible extraer un corpus de mayor volumen debido a la duración y a los recursos que son requeridos para llevar a cabo cada sesión del experimento.

Una rápida mejora sería contar con más de un ambiente silencioso y más personas para poder ejecutar el experimento en simultáneo a varios sujetos. De esta manera, se obtendrían más datos sin hacer cambios en la implementación. Otra mejora sería reemplazar al Mago de Oz por un reconocedor automático (ASR). Esto implicaría el beneficio de no requerir una persona que traduzca lo que dicen los sujetos. Sería necesario que el reconocedor distinga exactamente lo dicho, y el acto de diálogo ocurrido. Esto último se debe a que el Mago de Oz también se encarga del cumplimiento de las reglas por parte del sujeto. Este cambio permitiría recolectar datos en simultáneo al igual que la propuesta anterior. Otra mejora para aumentar el corpus, teniendo integrado el ASR, es que el juego se pueda jugar online. De esta forma, sería posible obtener un número mayor de muestras en un corto plazo.

Según estudios anteriores (Levitan et al. [2012] y Xia and Ma [2016]) se observan diferencias en la política de mimetización de acuerdo al género de los interlocutores. Es más notorio y frecuente este efecto entre pares mixtos de género. Debido a esto, nos pareció interesante ver qué ocurría en este experimento separando las elecciones por género de los sujetos. En este proyecto sólo se dan casos de pares mujer-mujer y hombre-mujer, ya que todos los avatares son de género femenino. Según los resultados de la Figura 24 las mujeres no presentan una preferencia por alguna política de mimetización, mientras que los hombres muestran una clara inclinación a elegir avatares con habla con entrainment. Sin embargo, no podemos llegar a una afirmación debido a que sólo hay 24 elecciones hechas por sujetos de cada género y este número es bajo.

Un cambio para realizar en el futuro sería que existan tanto avatares femeninos como masculinos. De esta manera, sería interesante observar qué efecto produce la voz del interlocutor en el sujeto. ¿Cambiarían las elecciones de los sujetos mujeres? ¿Y las de los hombres? ¿Influiría de la misma manera en la política de mimetización?

Se encontró que 32 de las elecciones cumplen con una condición. La misma consiste en elegir al avatar para jugar la partida final según el resultado de la última de la ronda. En otras palabras, si el sujeto gana la última partida elige al avatar correspondiente al mismo; si no, opta por el otro. Que se haya cumplido esta regla no implica que los sujetos hayan elegido según el resultado. Sin embargo, no se puede pasar por alto la inclinación observada. Durante el diseño y desarrollo de la aplicación no se tuvo en cuenta que el resultado de la última partida podría influir en las decisiones de los sujetos, por lo que, para descartar que los sujetos son afectados por el resultado de la partida habría que cambiar

una parte de la implementación. La propuesta sería integrar una variante más a los modos de juego, modificar el orden de los resultados de las partidas en cada ronda, el cual ahora es fijo.

En esta tesis se logró diseñar y desarrollar un sistema de diálogo hablado a través de un juego. El corpus obtenido ayudó a encontrar ciertas tendencias que pueden ser profundizadas en el futuro para llegar a resultados estadísticamente significativos, y por lo tanto, a conclusiones más fuertes. Además, se puede aprovechar el corpus generado en esta experimentación para estudiar otros aspectos y comportamientos de la comunicación entre un humano y la computadora a través de actos de diálogo como las preguntas y las respuestas.

A. Secuencias de las rondas

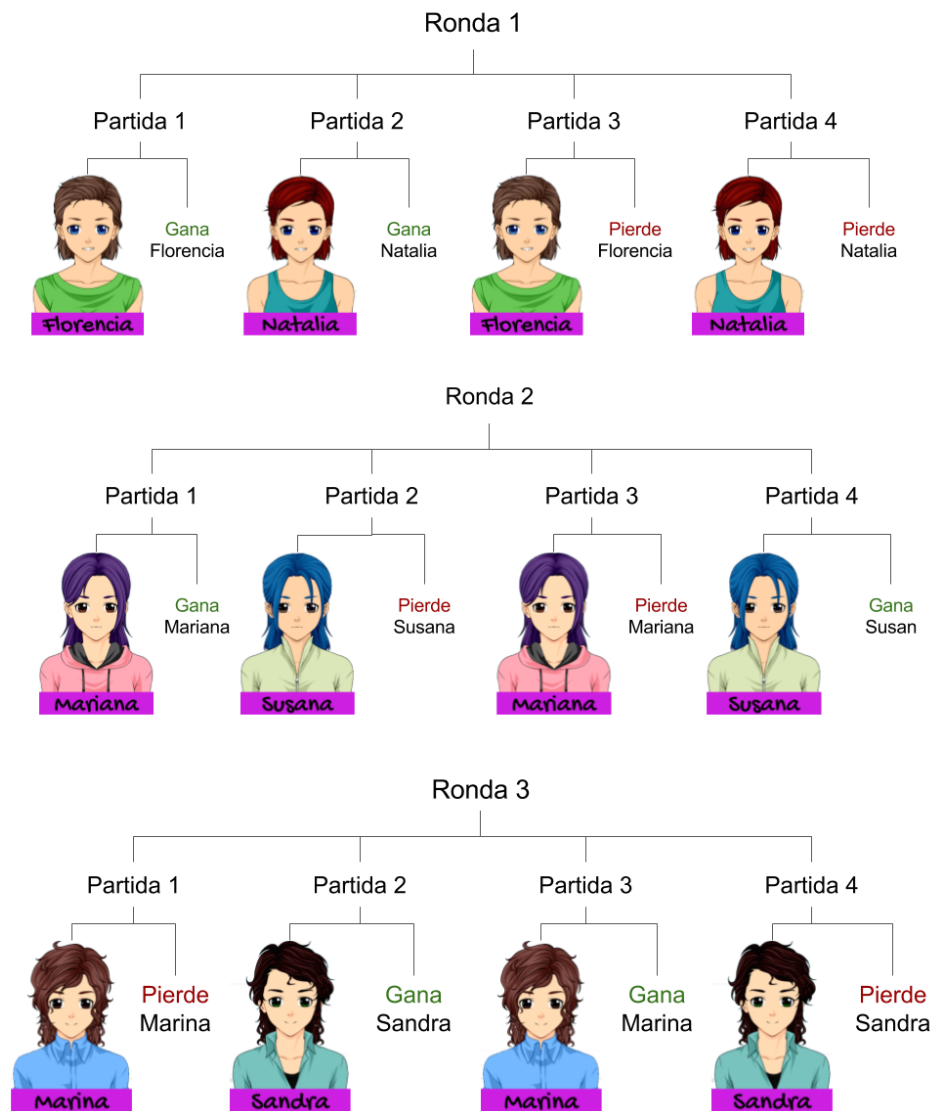


Figura 31: Configuración de las rondas 1, 2 y 3 para los modos de juego 2 y 4

B. Tableros de las rondas



Figura 32: Tablero correspondiente a la ronda 1

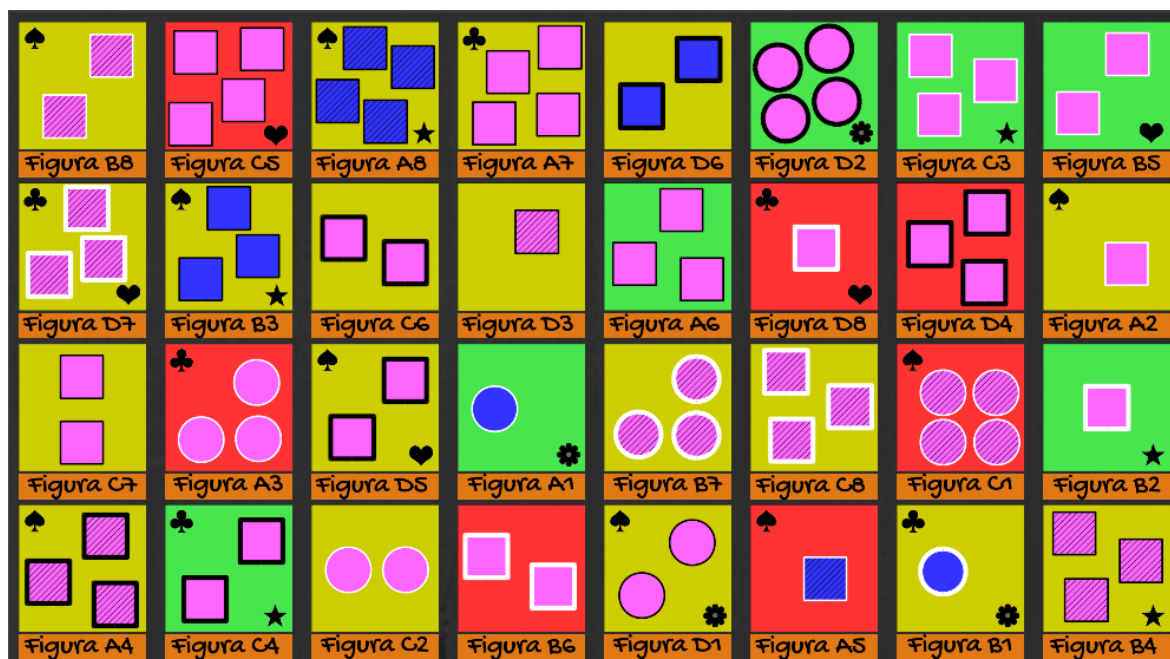


Figura 33: Tablero correspondiente a la ronda 2

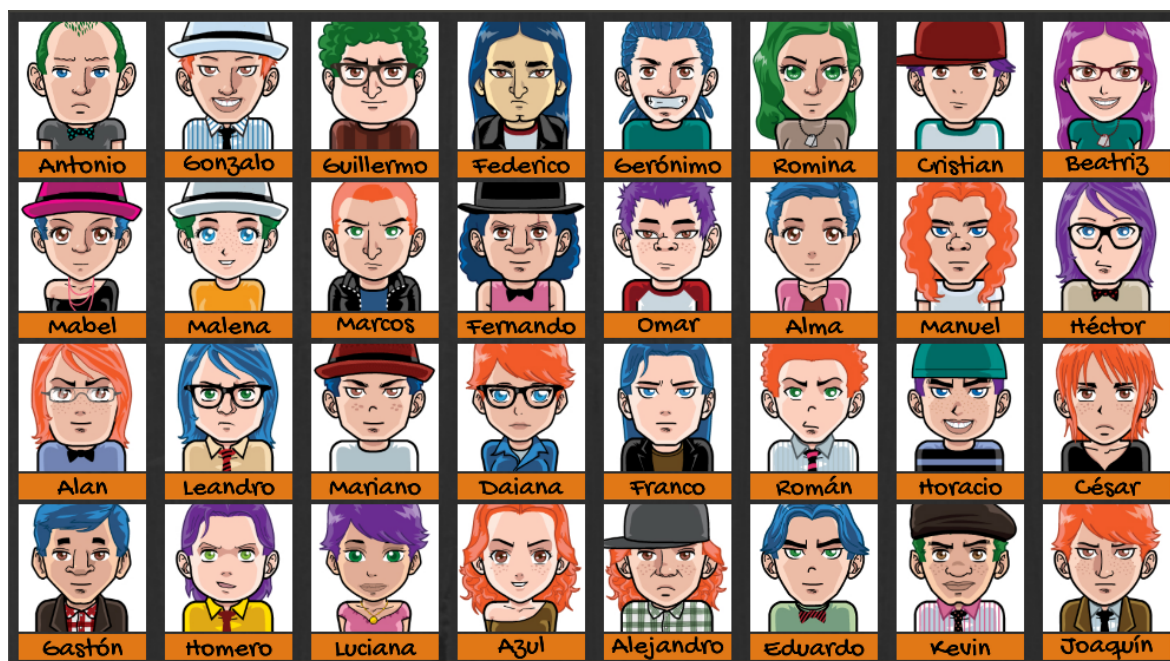


Figura 34: Tablero correspondiente a la ronda 3

C. Experimento

C.1. Cheat Sheet

Cheat Sheet

En el botón de audio encontrarás una etiqueta, que te va a ir guiando sobre lo que debés hacer.

Etiqueta	Descripción
PREGUNTAR	Es el turno de preguntarle a tu oponente.
RESPONDER	Deberás responder lo que te haya consultado el oponente.
ARRIESGAR	Te corresponde arriesgarte por alguna de las cartas que te quedaron levantadas en el tablero.
REPETIR	Deberá repetir lo dicho anteriormente o reformular tu pregunta con una propiedad válida. ¡Acordate de armar oraciones completas!
RESPONDER SI ACERTÓ	El oponente se arriesgó por una carta y deberás responder si es la misma que tenés.
ELEGIR Oponente	Luego de la cuarta partida en cada ronda, deberás decir contra que oponente querés jugar la siguiente partida.

- Ejemplos de pregunta:
 - ✓ ¿El color de pelo del personaje de la carta que debo adivinar es marrón?
 - ✓ ¿La persona de tu carta tiene puesto un collar?
 - ✓ ¿El fondo de las figuras de tu carta es de color azul?
 - ✓ ¿Tienen borde de color blanco las figuras de la carta?
 - ✓ ¿El personaje de la carta que debo adivinar tiene boca grande?
- Ejemplos de respuesta:
 - ✓ No, el personaje no tiene color de pelo marrón.
 - ✓ Sí, la carta tiene figuras con el borde grueso.

C.2. Texto introductorio

Cuando el sujeto ingresa al laboratorio, se le dice lo siguiente:

“Hola, antes que nada muchas gracias por participar de este experimento. Somos Agustín, Lara, Marisol y Ramiro y estamos llevando adelante este experimento dentro del marco de las actividades del Laboratorio de Inteligencia Artificial Aplicada del Departamento de Computación.

En este experimento van a jugar una serie de partidas del juego ¿Quién es Quién? contra avatares en la computadora, más adelante les vamos a explicar en detalle en qué consiste el juego. Pero queremos adelantarles que durante este experimento van a interactuar verbalmente con los mismos. Para nosotros es importante que cuando hablen con la computadora lo hagan de manera natural.

El experimento va a durar aproximadamente una hora. A lo largo del mismo pueden preguntarnos lo que quieran y, si por algún motivo lo desean, pueden dejar de participar cuando quieran. También es importante que sepan que todo lo que ustedes hagan en el mismo es totalmente anónimo y confidencial. En ningún momento registramos su nombre en ninguno de nuestros registros de datos.

Antes de comenzar el experimento les vamos a pedir que lean y firmen un consentimiento informado, después les vamos a pedir que completen una serie de cuestionarios para pasar luego a realizar las tareas en la computadora. A lo largo de la sesión les vamos a ir indicando qué hacer.

Si necesitaran ir al baño, jسته es el mejor momento! Por motivos técnicos necesitamos que apaguen sus celulares.

De vuelta muchas gracias por participar.”

Luego de que el sujeto complete los formularios, se lo dirige a la computadora. Cuando el sujeto haya finalizado la ronda de práctica, se le da la siguiente explicación:

“Muy bien. ¿Tenés alguna duda? Ahora vas a jugar la primera tanda de juegos. Recordá que vas a jugar 4 partidas (2 contra cada avatar, en algún orden) y luego, tenés que elegir un avatar para el juego final, que se juega por plata.”

C.3. Consentimiento



Universidad de Buenos Aires Formulario de Consentimiento

Investigador Principal: Agustín Gravano, Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Título del Protocolo IRB: Relación entre la confianza y el entrainment en el habla

Duración de la participación: 60 minutos

Número anticipado de sujetos: 20

Período: 01/09/2014 – 31/08/2017

Contacto: Agustín Gravano, gravano@dc.uba.ar, (+54) 11 5285-7478, Departamento de Computación, Oficina 15, Pabellón I, Ciudad Universitaria, Ciudad Autónoma de Buenos Aires.

Propósito de la Investigación

Estamos estudiando cómo las personas se comunican bajo diferentes condiciones.

Procedimientos de la Investigación

Vamos a solicitarle que juegue un juego de computadora en el laboratorio. Primero, habrá una sesión de práctica para que usted se acostumbre al diseño del juego. Luego usted jugará el juego e interactuará con voces de computadora (avatares). A continuación, usted completará un cuestionario acerca de la interacción; la evaluación incluirá algunas preguntas acerca de usted. Al final, usted recibirá información detallada acerca de la naturaleza del estudio y podrá realizar preguntas. Las grabaciones que haremos durante el experimento serán solamente de su voz, y las registraremos en la computadora. No registraremos ningún detalle identificatorio suyo. Solamente las personas involucradas en la realización del experimento tendrán acceso a las grabaciones, las cuales quedarán grabadas en un servidor seguro. Conservaremos estas grabaciones en forma indefinida, exclusivamente con fines de investigación.

Beneficios

No habrá beneficios por la participación en este experimento más allá de una compensación financiera.

Riesgos

Los riesgos de este estudio son similares a los que usted experimentaría al jugar un juego de computadora, tales como fatiga y cansancio ocular.

Confidencialidad

En este experimento, usted será identificado mediante un número de sujeto, el cual no estará vinculado a su nombre de ninguna manera. No registraremos su nombre en ninguno de nuestros registros de datos. Su formulario de consentimiento será guardado en archivos separados para registrar que usted participó en este experimento. Si alguna publicación o presentación resultare de esta investigación, usted no será identificado por el nombre.

Participación Voluntaria

La participación en este estudio es voluntaria. La negativa a participar no conllevará penalidad alguna o pérdida de beneficios que de otro modo le correspondieran. Del mismo modo, si usted elige participar en este estudio, puede interrumpir su participación en cualquier momento, sin que esto conlleve penalidad alguna o pérdida de beneficios que de otro modo le correspondieran.

Edad

Al firmar este formulario de consentimiento usted certifica que tiene entre 18 y 65 años de edad.

Compensación

Usted recibirá una paga a razón de 90 pesos argentinos por hora (por ejemplo, si le lleva 45 minutos completar el estudio, usted recibirá 67.50 pesos argentinos). Puede ganar un bonus adicional de hasta 90 pesos argentinos, en base a su performance en el juego. No recibirá compensación adicional por permitir que su voz sea grabada.

Información Adicional

Si usted tiene preguntas o resulta lastimado mientras toma parte de este estudio de investigación, debe contactar a Agustín Gravano, gravano@dc.uba.ar, (+54) 11 5285-7478, Departamento de Computación, Oficina 15, Pabellón I, Ciudad Universitaria, Ciudad Autónoma de Buenos Aires.

Si usted tiene preguntas acerca de sus derechos como un sujeto de investigación, debe contactar al Comité de Ética del Centro de Educación Médica e Investigaciones Clínicas (CEMIC), pechavarria@cemic.edu.ar, (+54) 11 5299-0100, int. 2879, Av. Galván 4102, Ciudad Autónoma de Buenos Aires.

Firmas

Sujeto del Estudio

Aclaración_____Firma_____Fecha_____

Persona que Obtiene el Consentimiento

Aclaración_____Firma_____Fecha_____

C.4. Formulario de personalidad

PARTICIPANTE #: _____

Por favor, respondé las siguientes preguntas usando esta escala de valores:

1	2	3	4	5	6	7
Desacuerdo fuerte	Desacuerdo moderado	Desacuerdo leve	Ni acuerdo ni desacuerdo	Acuerdo leve	Acuerdo moderado	Acuerdo fuerte

“Me veo a mí mismo como...”

1. Extrovertido, entusiasta.	1: <input type="checkbox"/>	2: <input type="checkbox"/>	3: <input type="checkbox"/>	4: <input type="checkbox"/>	5: <input type="checkbox"/>	6: <input type="checkbox"/>	7: <input type="checkbox"/>
2. Crítico, peleador.	1: <input type="checkbox"/>	2: <input type="checkbox"/>	3: <input type="checkbox"/>	4: <input type="checkbox"/>	5: <input type="checkbox"/>	6: <input type="checkbox"/>	7: <input type="checkbox"/>
3. Confiante, controlado.	1: <input type="checkbox"/>	2: <input type="checkbox"/>	3: <input type="checkbox"/>	4: <input type="checkbox"/>	5: <input type="checkbox"/>	6: <input type="checkbox"/>	7: <input type="checkbox"/>
4. Nervioso, que se altera fácil.	1: <input type="checkbox"/>	2: <input type="checkbox"/>	3: <input type="checkbox"/>	4: <input type="checkbox"/>	5: <input type="checkbox"/>	6: <input type="checkbox"/>	7: <input type="checkbox"/>
5. Abierto a nuevas experiencias, con múltiples intereses.	1: <input type="checkbox"/>	2: <input type="checkbox"/>	3: <input type="checkbox"/>	4: <input type="checkbox"/>	5: <input type="checkbox"/>	6: <input type="checkbox"/>	7: <input type="checkbox"/>
6. Reservado, callado.	1: <input type="checkbox"/>	2: <input type="checkbox"/>	3: <input type="checkbox"/>	4: <input type="checkbox"/>	5: <input type="checkbox"/>	6: <input type="checkbox"/>	7: <input type="checkbox"/>
7. Comprensivo, afectuoso.	1: <input type="checkbox"/>	2: <input type="checkbox"/>	3: <input type="checkbox"/>	4: <input type="checkbox"/>	5: <input type="checkbox"/>	6: <input type="checkbox"/>	7: <input type="checkbox"/>
8. Desorganizado, descuidado.	1: <input type="checkbox"/>	2: <input type="checkbox"/>	3: <input type="checkbox"/>	4: <input type="checkbox"/>	5: <input type="checkbox"/>	6: <input type="checkbox"/>	7: <input type="checkbox"/>
9. Calmo, emocionalmente estable.	1: <input type="checkbox"/>	2: <input type="checkbox"/>	3: <input type="checkbox"/>	4: <input type="checkbox"/>	5: <input type="checkbox"/>	6: <input type="checkbox"/>	7: <input type="checkbox"/>
10. Convencional, poco creativo.	1: <input type="checkbox"/>	2: <input type="checkbox"/>	3: <input type="checkbox"/>	4: <input type="checkbox"/>	5: <input type="checkbox"/>	6: <input type="checkbox"/>	7: <input type="checkbox"/>

C.5. Instrucciones del juego

¿Quién es Quién?

El juego que vas a jugar se llama “**¿Quién es Quién?**”. Vos vas a ser uno de los jugadores y tus contrincantes serán personajes que usan distintos programas de inteligencia artificial.

El juego empieza con una ronda de entrenamiento para que te familiarices. Luego, habrá tres rondas más. En cada ronda, jugarás contra dos personajes distintos. Además, contra cada uno de ellos jugarás dos veces por ronda para poder ver el nivel de cada uno. Luego, vos **deberás elegir contra cuál de los dos oponentes querés jugar en el juego final**. Ese juego será **por plata**, de modo que tenés que elegir el que peor juege.

Cada participante jugará con su propio tablero que tiene 32 cartas distintas. Ambos tableros tienen las mismas fichas ordenadas de manera diferente.

El **objetivo** del juego es intentar adivinar el personaje de la carta elegida para el oponente. A su vez, el contrincante intentará hacer lo mismo con tu carta.

Para lograr el objetivo del juego se deberá ir haciendo preguntas en cada turno por alguna característica permitida. Éstas estarán listadas debajo de tu tablero. Antes de formular tu pregunta, fijate que esté la característica por la que querés averiguar.

De acuerdo a la respuesta, podrás ir bajando fichas de tu tablero, para ir descartando y lograr adivinar la carta que tiene el oponente. Luego que el oponente responda la consulta que has hecho, te aparecerá un cartel indicando que se bajarán las cartas que cumplan esa característica. Deberás **aceptar** para que se bajen.

Así, luego de varios turnos, a vos o tu oponente les quedará una sola ficha boca arriba en el tablero. Si te ocurrió a vos, el contrincante deberá arriesgarse por una carta de las de su tablero, ya que en el siguiente turno vos ya sabrás la ficha que tiene. Si acertó, **perdiste**. Caso contrario, **¡ganaste!** 😊

Lo mismo ocurre si a tu oponente le queda una sola ficha en el tablero. En tu turno deberás arriesgarte por una carta de todas las que te quedaron “levantadas” en tu tablero. Si acertaste, **¡ganaste!** Caso contrario, **perdiste**. 😞

La comunicación con tu contrincante será a través de un botón de audio. Para hablar deberás mantener presionado el botón con el mouse. Al terminar de hablar, **mantené apretado unos segundos** y luego soltalo (al estilo *Whatsapp*).



Reglas

- ¡No se puede mentir!



- No podés arriesgarte por una ficha a menos que a tu oponente le quede una sola en el tablero.
- En todas las partidas empezarás vos.
- Para formular tus preguntas, en cada turno, sólo podrás usar cierto conjunto de características. Si tu consulta es sobre una propiedad inválida, el juego no te permitirá continuar hasta que indiques una permitida.

A partir del cuarto turno de cada partida se habilitará una característica nueva. Prestá atención al listado ya que te ayudará bastante.

- Ejemplos de preguntas que puedes hacer:
 - ✓ ¿El color de pelo del personaje de la carta que debo adivinar es marrón?
 - ✓ ¿La persona de tu carta tiene puesto un collar?
 - ✓ ¿El fondo de las figuras de tu carta es de color azul?
 - ✓ ¿Tienen borde de color blanco las figuras de la carta?
 - ✓ ¿El personaje de la carta que debo adivinar tiene boca grande?
- Por otro lado, estos son ejemplos de consultas mal estructuradas:
 - ✗ ¿Lleva sombrero?
 - ✗ ¿Hay cuadrados?
- Tus preguntas deberán permitir responder únicamente de forma afirmativa o negativa. Para las respuestas, no se podrá contestar sólo con “**Sí**” o “**No**”, sino que además necesitamos que repitas la característica por la cual te hicieron la consulta. Por ejemplo:
 - ✓ No, el personaje no tiene color de pelo marrón.
 - ✓ Sí, la carta tiene figuras con el borde grueso.

Te pedimos que por favor armes oraciones completas.

Te pedimos, por favor, que luego de jugar la ronda de prueba, avises que la finalizaste antes de continuar.

¡Gracias!

Cualquier duda, preguntanos.

¡A jugar!

✿ ¡Mucha Suerte! ✿

C.6. Formulario de formación

PARTICIPANTE #: _____

Tus respuestas a estas preguntas sólo buscan proveer información general de los participantes de nuestro estudio, y de ninguna manera afectan a tu participación en el mismo. Toda la información se mantendrá en forma estrictamente **confidencial** y **anónima**, y tu nombre **no** se adjuntará a la misma.

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 idiomas en forma fluida? Sí: ☐ No: ☐

En caso afirmativo, listá los idiomas que hablás (incluido 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).

a. _____ ESC: ☐ ENT : ☐ (desde la edad: _____)b. _____ ESC: ☐ ENT : ☐ (desde la edad: _____)c. _____ 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: _____)

C.7. Recibo



RECIBO

Recibí del Laboratorio de Inteligencia Artificial Aplicada (Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires) la cantidad de \$_____ (pesos argentinos _____) en concepto de viáticos por mi participación como voluntario/a en el estudio correspondiente al Proyecto de Investigación “Relación entre la confianza y el entrainment en el habla”.

Son \$_____

Fecha: / / 201__

Firma:

Aclaración:

DNI:

C.8. Detalles del experimento



Universidad de Buenos Aires

Declaración de Parte (*Debriefing Form*)

Investigador Principal: Agustín Gravano, Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Título del Protocolo IRB: Relación entre la confianza y el entrainment en el habla

¡Gracias por aceptar participar en este estudio! El propósito general de este proyecto es investigar si la gente confía más en avatares cuya habla se asemeja más a la suya.

Usted ha jugado un juego de computadora e interactuado con avatares durante el curso del juego. Nosotros hemos manipulado sus voces para que fueran similares o distintas a la suya. Basándonos en sus selecciones de consejos de los avatares y las respuestas al cuestionario post-estudio, estudiaremos si la manipulación de las voces de los avatares afectaron sus elecciones y la confianza general hacia los avatares. Los resultados de este estudio pueden llevar a interacciones verbales humano-computadora que sean más naturales y efectivas.

Por favor, maneje esta información de manera confidencial y no discuta los procedimientos de este estudio con otros posibles sujetos.

Gracias nuevamente por su participación en el estudio. Usted todavía puede revocar su consentimiento de uso de las grabaciones de audio, así como renunciar al estudio por completo. Una vez se haya retirado del laboratorio, no podremos acceder a un posible pedido de renuncia, dado que no será posible identificar sus datos.

Si usted tiene más preguntas sobre este estudio, por favor contacte a Agustín Gravano, gravano@dc.uba.ar, (+54) 11 4576-3390 int. 717, Departamento de Computación, Oficina 15, Pabellón I, Ciudad Universitaria, Ciudad Autónoma de Buenos Aires. Además, si usted tiene inquietudes sobre cualquier aspecto de este estudio, debe contactar al Comité de Ética del Centro de Educación Médica e Investigaciones Clínicas (CEMIC), pechavarria@cemic.edu.ar, (+54) 11 5299-0100, int. 2879, Av. Galván 4102, Ciudad Autónoma de Buenos Aires.

D. Resultados complementarios

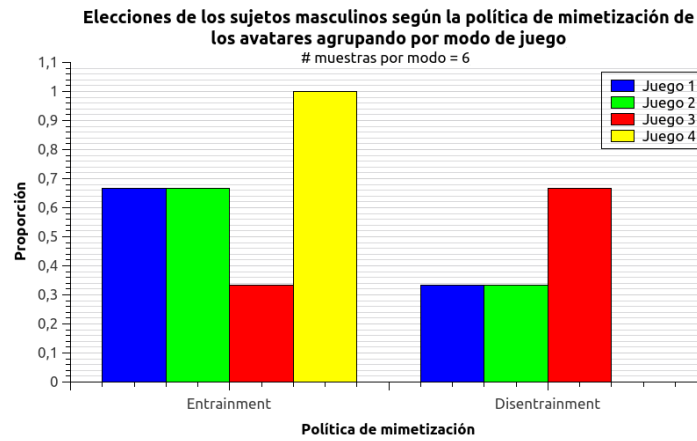


Figura 35: Proporción de los hombres que eligieron *entrainment* contra *disentrainment* por modo de juego

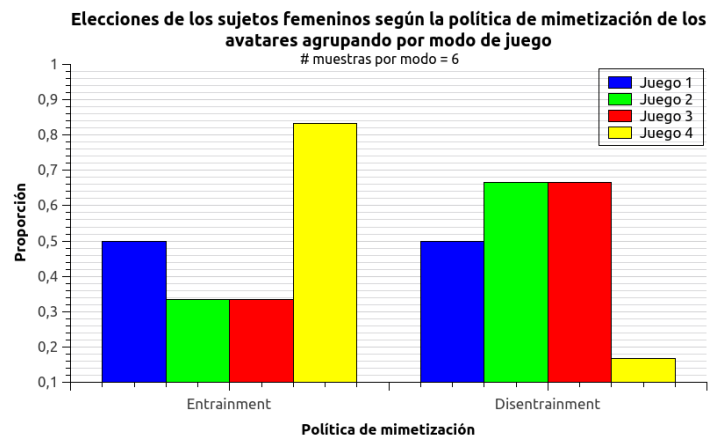


Figura 36: Proporción de las mujeres que eligieron *entrainment* contra *disentrainment* por modo de juego

Referencias

- S. E. Brennan and H. H. Clark. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6):1482, 1996.
- T. L. Chartrand and J. A. Bargh. The chameleon effect: the perception-behavior link and social interaction. *Journal of personality and social psychology*, 76(6):893, 1999.
- Z. Hernández-Figueroa, F. J. Carreras-Riudavets, and G. Rodríguez-Rodríguez. Automatic syllabification for spanish using lemmatization and derivation to solve the prefix’s prominence issue. *Expert Systems with Applications*, 40(17):7122–7131, 2013.
- D. Jurafsky, E. Shriberg, B. Fox, and T. Curl. Lexical, prosodic, and syntactic cues for dialog acts. In *Proceedings of ACL/COLING-98 Workshop on Discourse Relations and Discourse Markers*, pages 114–120, 1998.
- R. Levitan and J. Hirschberg. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Twelfth Annual Conference of the International Speech Communication Association*, 2011.
- R. Levitan, A. Gravano, L. Willson, S. Benus, J. Hirschberg, and A. Nenkova. Acoustic-prosodic entrainment and social behavior. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies*, pages 11–19. Association for Computational Linguistics, 2012.
- M. Natale. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, 32(5):790, 1975.
- J. S. Pardo. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4):2382–2393, 2006.
- J. M. Pérez, R. H. Gálvez, and A. Gravano. Disentrainment may be a positive thing: A novel measure of unsigned acoustic-prosodic synchrony, and its relation to speaker engagement. In *INTERSPEECH*, pages 1270–1274, 2016.
- D. Reitter and J. D. Moore. Predicting success in dialogue. 2007.
- G. Skantze and A. Hjalmarsson. Towards incremental speech generation in dialogue systems. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 1–8. Association for Computational Linguistics, 2010.
- M. Ter Maat, K. P. Truong, and D. Heylen. How turn-taking strategies influence users’ impressions of an agent. In *IVA*, volume 6356, pages 441–453. Springer, 2010.
- L. Violante, P. R. Zivic, and A. Gravano. Improving speech synthesis quality by reducing pitch peaks in the source recordings. In *HLT-NAACL*, pages 502–506, 2013.
- Z. Xia and Q. W. Ma. Gender and prosodic entrainment in mandarin conversations. In *Chinese Spoken Language Processing (ISCSLP), 2016 10th International Symposium on*, pages 1–4. IEEE, 2016.