

# Relación entre el gesto y la voz durante la percepción de enunciados declarativos e interrogativos del español

## Relationship between Gesture and Voice in the Perception of Spanish Declarative and Interrogative Utterances

**Hernán Martínez**

*Laboratorio de Fonética  
Universidad de Los Andes  
Mérida, Venezuela  
[hernan.martinez@ula.ve](mailto:hernan.martinez@ula.ve)*

**Elsa Mora**

*Laboratorio de Fonética  
Universidad de Los Andes  
Mérida, Venezuela  
[elsamora@ula.ve](mailto:elsamora@ula.ve)*

### Resumen

En la onda sonora de interrogativas y declarativas está codificada la distinción acústica entre estos dos tipos de enunciados. En este trabajo nos centraremos en el hecho de querer determinar hasta qué punto los oyentes hacen uso de la información visual para establecer la modalidad sintáctica a la que pertenecen los enunciados que perciben. Consideramos que el acceso a la información visual durante la percepción del habla influye sobre el resultado final del proceso de percepción de este tipo de enunciados. Luego de la aplicación de dos tests (uno auditivo y otro audiovisual) afirmamos que el acceso a la información visual durante la percepción de los enunciados interrogativos y declarativos influye en el resultado final de este proceso.

**Palabras clave:** Percepción del habla, gestualidad, enunciados interrogativos, enunciados declarativos, prosodia.

### Abstract

Acoustic distinction of interrogative and declarative utterances is coded in their sound waves. This study has the objective of determining the influence visual information has in listeners for establishing grammar category when listening to utterances. Access to visual information during speech perception influences on final perception processing. An auditory perception test and an audiovisual perception test were applied. Results show that visual information during speech perception has an influence on listeners' perception.

**Keywords:** Speech perception, gestures, interrogative utterance, declarative utterance, prosody.

## 1. INTRODUCCIÓN

Los seres humanos están equipados con múltiples canales sensoriales perceptores para percibir que de forma conjunta proporcionan información proveniente del exterior. Por lo tanto, es un reto para el cerebro dar sentido a estas múltiples fuentes de información, unir las señales procedentes de ellas y distinguirlas de otras señales.

El proceso de percepción del habla se ha vislumbrado bastante complejo en varios de los estudios realizados en el ámbito lingüístico (cf. Maltin y Foley, 1996; Yeni-Komshian, 1999; Marrero, 2001 y Martínez, 2009). Estos estudios han explicado cómo las personas son capaces de percibir los límites entre las palabras y también los límites entre elementos menores como la sílaba o los segmentos fónicos (aun cuando estos carecen de pausas y se caracterizan por ser de naturaleza variable). Esta realidad experimental ha sido fundamental para el desarrollo de esos trabajos y de varios modelos de percepción del habla, ya que con ellos se evidencia que el proceso de percepción se debe a procesos de segmentación e interpretación realizados en niveles superiores de procesamiento cognitivo que involucran la activación de varios niveles del sistema lingüístico. Así, según Martínez (2009:35), la percepción del habla es “un proceso global, estructurado y orientado temporalmente, que involucra la sensación y la experiencia o el conocimiento previo del sujeto e implica la segmentación de las unidades lingüísticas y la interpretación o asignación de un significado”.

Con base en el hecho de que la señal acústica es rica en información sobre el lenguaje hablado, muchas de las investigaciones sobre percepción del habla se han centrado exclusivamente en ella, dejando de lado la señal visual que, tal como lo señalan los estudios provenientes de la psicología experimental, puede tener un gran impacto en la percepción de la señal auditiva bajo condiciones degradantes. Así, se ha demostrado que la disponibilidad de información visual modifica la percepción de la información acústica del habla; que la experiencia perceptiva es diferente según se disponga de uno (auditivo) o dos canales de información (audiovisual). Una de las evidencias de este hecho es el efecto o ilusión McGurk (MacDonald y McGurk, 1978), nombre con el que se conoce a la demostración del fenómeno audiovisual en la percepción del habla.

Los resultados del trabajo de MacDonald y McGurk (1978) demuestran que si la sílaba auditiva [ba] se muestra junto a un video de un hablante que pronuncia la sílaba [ga], el resultado perceptivo es [da]; una sílaba nueva e inexistente en las modalidades sonora y visual por separado. La percepción visual de los movimientos articulatorios del hablante modifica, de manera significativa, la experiencia perceptiva de la escucha. Tal como sostienen Martínez *et al.* (2012:167), “se podría afirmar que las señales visuales y auditivas dispares generan un efecto subjetivo: una ilusión de ventrilocuismo. La integración de esas dos señales es una propiedad del sistema perceptivo de los humanos quienes en gran variedad de situaciones, forman lo que MacDonald y McGurk llaman *unidades de creencia audio-visuales*”.

Varias investigaciones (cf. por ejemplo, Jaekl *et al.* 2011) han comprobado que la integración audiovisual del habla es un fenómeno sólido e inmune a una variedad de situaciones experimentales, como la experimentación en ambientes ruidosos. Este

fenómeno es sólido, por ejemplo, durante la percepción de lengua materna o de una segunda lengua; también durante la presentación de información sonora con un tiempo de retardo severo respecto de la información visual. Se ha podido demostrar que también es sólido en casos en los que la señal acústica se presenta manipulada o intacta (cf. por ejemplo, Martínez *et al.*, 2012; Borrás *et al.*, 2013; Crespo *et al.*, 2013); o si la presentación de la información acústica y visual no se corresponde con un hablante del mismo género (Borrás *et al.*, 2013). También se ha demostrado la solidez de la integración audiovisual durante la percepción del habla en situaciones en las que las informaciones sonoras y visuales provienen de fuentes naturales o animadas (Soto, 2000).

Massaro y Cohen (1983), a través de un experimento mediante el cual se proponían describir la manera en que la información auditiva y la visual se integran en el habla, han reportado que a pesar de haber instruido expresamente a los sujetos del experimento para que identificaran sólo lo que escuchaban, la percepción visual de las articulaciones del hablante había modificado considerablemente el resultado perceptivo sonoro. Massaro y Cohen también afirman que los movimientos articulatorios visibles durante la percepción del habla aportan inteligibilidad a la decodificación del mensaje sonoro en situaciones de degradación acústica. Según estos investigadores, la percepción de los movimientos de la boca es potencialmente suficiente para identificar ciertos aspectos del habla en sujetos con capacidades auditivas normales y en condiciones acústicas adecuadas porque, en muchos casos, los segmentos fonéticos tienen equivalentes en los movimientos de la boca. En este trabajo de Massaro y Cohen se concluye que la integración audiovisual es el resultado de una tarea en la que el sujeto evalúa las informaciones provenientes de cada fuente. El producto perceptivo del habla bimodal es, por tanto, una sinergia del procesamiento de las informaciones sonora y visual. Este hecho también lo afirman Bertelson *et al.* (2000) y Soto (2000), al decir que los sistemas visual y auditivo se soportan entre sí y la información general obtenida se integra en el proceso de percepción.

Afirmaciones como estas encuentran fundamento también en trabajos como el de Sekiyama *et al.* (2003). A través del uso de la resonancia magnética funcional y de la tomografía por emisión de positrones, estos investigadores han estudiado el comportamiento del cerebro durante la percepción auditiva, visual y audiovisual del lenguaje. Los datos obtenidos a partir de las herramientas imagenológicas revelan que hay activación cerebral bilateral específica durante la percepción de estímulos auditivos en la corteza temporal y activación en el MT /V5 durante la percepción visual. Ahora bien, durante la percepción de los estímulos audiovisuales se pudo observar que se produce activación en la corteza temporal izquierda que se extiende hacia el área visual posterior superior del cerebro, sobre todo cuando son presentados estímulos incongruentes. También encontraron los investigadores que hay un aumento de la actividad cerebral en la parte posterior del surco temporal izquierdo, lo que indicaría que esta región del cerebro participa fuertemente durante el procesamiento y unificación de la información audiovisual. Estos datos también han sido reportados luego de estudios con Resonancia Magnética Funcional por Latinus *et al.* (2011).

A partir de todas estas investigaciones se deduce, por una parte, que las condiciones en que se presenta la información audiovisual ejerce una potente influencia sobre el resultado perceptivo del habla, y por otra, que el habla bimodal debe ser observada como una

coalición, una alianza de componentes. Los trabajos sobre la integración audiovisual advierten de la naturaleza del efecto que la integración audiovisual del habla produce en los perceptores: el resultado reportado del habla audiovisual es una poderosa unidad de creencia formada por la combinación de los canales auditivo y visual que se resiste a la identificación fragmentada o parcial de sus componentes, aunque se insista en inducirla.

### **1.1 Gesto y voz en la percepción de declarativas e interrogativas**

Durante los últimos 20 años se han desarrollado investigaciones que se han centrado en el estudio del procesamiento tanto de entrada (*input*) como de salida (*output*) de señales simultáneas provenientes de varias fuentes durante la producción y percepción del lenguaje. Uno de los casos más estudiados es el que analiza la relación entre el gesto manual o facial y la voz. Pietrosemoli y Mora (1998), por ejemplo, sostienen que durante la producción del lenguaje existe una relación muy estrecha entre la gestualidad y la entonación que se manifiesta en una sincronización entre la elevación y mantenimiento de la F0 y la elevación de la mano. Esteve-Gibert y Prieto (2013), por su parte, a través de una experiencia perceptiva con catalanes, sostienen la idea de que existe una regla de sincronización fonológica que muestra que la aparición de gestos se da junto con la aparición del pico entonativo de la frase emitida.

Los usuarios de las lenguas, además de marcas sintácticas y entonativas, utilizan marcas gestuales para generar la diferencia entre frases declarativas e interrogativas. De hecho, las expresiones faciales, los movimientos de la cabeza y la mirada, parecen estar profundamente relacionados con la producción y la percepción del habla. Este hecho lo han señalado Borrás *et al.* (2013) en un estudio en el que revelan que tanto en la producción como en la percepción de frases interrogativas en holandés y catalán existen diferencias auditivas específicas que las hacen distintas de otro tipo de frase, pero también señalan que la mirada y el movimiento de las cejas son marcas gestuales que se utilizan para crear esta distinción. Este hecho también lo afirman para el catalán Crespo *et al.* (2013). Al analizar la producción de enunciados interrogativos en alemán y catalán, estos autores también afirman que en alemán las marcas entonacionales son más importantes que las gestuales. En todo caso, en estos dos trabajos se afirma que la percepción depende en gran medida de la información auditiva para determinar si un enunciado es una pregunta o una afirmación, pero la precisión se mejora aún más cuando se añade información visual.

Incluso, pareciera ser importante y determinante contar con la información visual durante la percepción de la doble negación en catalán (Prieto *et al.*, 2013) y durante la interpretación de las respuestas afirmativas a las preguntas negativas (yes/no questions) también en catalán (Tubau *et al.*, 2014).

La importancia del acceso a la información visual durante la percepción del habla se ha visto también en experimentos realizados en el ámbito de la síntesis de voz. Srinivasan & Massaro (2003) hicieron uso de representaciones sintéticas de un rostro humano en las que se variaban las características auditivas y visuales específicas con el fin de investigar si los perceptores podrían diferenciar entre frases declarativas e interrogativas en inglés. Los autores encontraron que tanto la elevación de las cejas como la inclinación de la cabeza podrían aumentar la detección de una pregunta por parte de los perceptores. Por su parte,

Ouni *et al.* (2006) señalan la importancia de los rasgos faciales visibles durante la percepción del habla emitida por sintetizadores animados. Los investigadores afirman que los sintetizadores analizados deben mejorar contundentemente los movimientos faciales relacionados con los sonidos del lenguaje y la expresión del mismo con el fin de lograr mejores resultados en la credibilidad y percepción de los mensajes emitidos por esos sintetizadores.

La influencia de la información visual durante la percepción de frases declarativas e interrogativas también fue estudiada en el español venezolano por Higuera (2013). En su trabajo, la autora sostiene que durante el proceso perceptivo la información visual juega un papel preponderante y decisivo. Afirma Higuera (2013) que los oyentes toman sus decisiones perceptivas mayormente con base en lo que ven y no tanto en lo que escuchan. Al igual que en otras investigaciones ya citadas, Higuera afirma que cuando los oyentes cuentan con un input visual y uno auditivo incongruentes, los oyentes presentan dificultades para percibir el mensaje, decidiendo en la mayoría de los casos por la información visual. Este hecho también ha sido señalado por Martínez *et al.* (2012), quienes señalan que los oyentes están cotejando siempre ambos inputs durante el proceso de percepción del habla. Estos autores confirman que la atención recae en aquel input que mejor se adapte a las expectativas (contextuales/situacionales, culturales, sociales, anímicas, etc.) del oyente. También señalan que los oyentes atienden mejor al input visual que al auditivo durante la percepción audiovisual.

Tanto Higuera (2013) como Martínez *et al.* (2012), quienes han trabajado con estímulos manipulados a los que se les ha sustituido el audio o el video, afirman que el input visual es importante a la hora de escoger ante estímulos manipulados o incongruentes: cuando se presentan este tipo de estímulos, en los que el input visual no se corresponde con el input auditivo, los oyentes seleccionan mayoritariamente la información visual.

Partiendo de la observación que hemos hecho de las grabaciones con las que hemos trabajado y de los datos obtenidos por otros investigadores proponemos para nuestro estudio un sistema binario que nos permita entender cuándo aparecen los gestos faciales y cuándo no, y en qué momento se combinan para transmitir una información relacionada con el tipo de enunciado emitido por los hablantes. Luego de la observación minuciosa de las producciones de los hablantes consideramos tres variables para establecer la siguiente matriz de rasgos con el fin de sistematizar la caracterización de los gestos faciales: movimiento del ceño, movimiento de las cejas, movimiento de la cabeza. Esta matriz de rasgos no pretende ser rigurosa, pero es suficiente para entender la configuración de los movimientos faciales usada por los hablantes al emitir los enunciados que estamos estudiando. Así, tendríamos en la tabla 1 los dos enunciados analizados en esta investigación y sus respectivos rasgos.

Tabla 1. Rasgos gestuales para cada uno de los enunciados. Los símbolos + y – señalan la presencia o la ausencia del rasgo, respectivamente.

	Declarativos	Interrogativos
<b>Fruncimiento del ceño</b>	-	+
<b>Elevación de las cejas</b>	+	-
<b>Inclinación de la cabeza</b>	-	+

El conjunto de estos rasgos gestuales presentes en la producción de enunciados interrogativos y declarativos vendría a constituir lo que Cavé (2003) llama *visemas* o lo que MacDonald y McGurk llaman *unidades de creencia audio-visuales*. Según Cavé (2003:63), “el efecto McGurk ha dado origen a una línea de investigación con miras a describir e identificar los “visemas”, equivalentes visuales de los fonemas” (cf. figura 1).

## 1.2. Distinción acústica entre enunciados declarativos e interrogativos

Ahora bien, aunque en las distintas lenguas del mundo la distinción entre enunciados declarativos e interrogativos puede estar marcada por modificaciones morfosintácticas y de entonación, estar marcada morfosintácticamente o estar marcada solo por la entonación (Frota *et al.*, 2013), parece probable la idea de la existencia de una tendencia universal a usar una F0 alta para marcar los enunciados interrogativos y una F0 baja para marcar los enunciados declarativos (Soderstrom *et al.*, 2011; Yang y Liang, 2012). Evidencia adicional la aportan Yuan *et al.* (2002), quienes sostienen que las diferencias entre declarativas e interrogativas en chino se deben a valores altos de la F0 en los tonos finales y a un rango tonal más alto en las oraciones interrogativas. Estos resultados, sumados a los del estudio de Gussenhoven y Chen (2000), indican que las personas tienden a percibir los enunciados con entonación alta y/o con tonos finales altos como interrogativos independientemente de la lengua materna. Sin embargo, Cangemi y D’Imperio (2013) señalan que las lenguas romance, germánicas y las no-indoeuropeas, además de mostrar contrastes claros entre los enunciados declarativos e interrogativos basados en las variaciones de la F0, marcan contrastes también a través de los cambios de la duración.

Estas afirmaciones se comprueban en los señalamientos hechos por otros autores para el español. Astruc *et al.* (2010:20) señalan que las interrogativas absolutas presentan sobre todo un tono más elevado que las declarativas. Méndez (2010:16), en su artículo con datos del proyecto AMPER, refiere que “un menor campo tonal interactúa con un mayor campo durativo en la configuración de frases declarativas; en las interrogativas sucede lo contrario: menor campo durativo y mayor campo tonal”. Finalmente, en un estudio comparativo con datos del proyecto AMPER de diferentes variedades del español de Canarias y diferentes variedades del español venezolano se concluye que “las pendientes Inicio-Final generadas por las interrogativas presentan mayor altura tonal que las declarativas en todas las variedades estudiadas” (Dorta *et al.*, 2013:215).

## 2. METODOLOGÍA

Partiendo del supuesto de que en la onda sonora de declarativas e interrogativas se encuentra codificada la distinción acústica entre estos dos tipos de enunciados, en este trabajo nos centraremos en el hecho de querer determinar hasta qué punto los oyentes hacen uso de la información visual para establecer la modalidad sintáctica a la que pertenecen los enunciados que perciben. Creemos que el acceso a la información visual durante la percepción del habla influye sobre el resultado final del proceso de percepción.

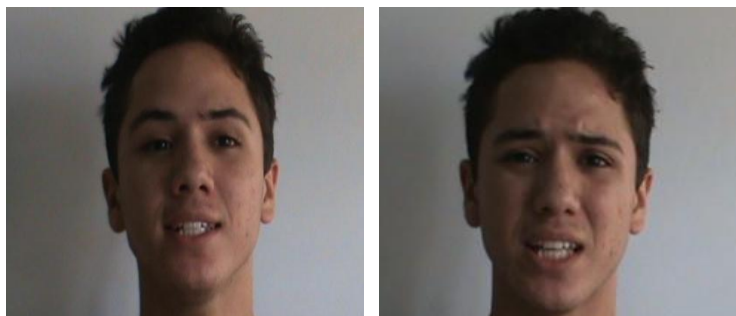
## 2.1. Las grabaciones

Con la intención de determinar si la expresión facial influye en la percepción de la expresión hablada de los enunciados declarativos e interrogativos, se trabajó con una parte de un corpus para el estudio de la influencia de la información visual durante la percepción del habla recogido por integrantes y colaboradores del Grupo de Investigación en Ciencias Fonéticas de la Universidad de Los Andes.

Específicamente trabajamos con un grupo constituido por 20 grabaciones audiovisuales: 10 de ellas contienen la emisión de enunciados declarativos y las otras 10 contienen los mismos enunciados pero emitidos de manera interrogativa. Así tenemos frases como *en Venezuela se habla español/ ¿En Venezuela se habla español? Su familia es de Mérida/ ¿su familia es de Mérida?*, entre otros. Los 20 enunciados fueron recogidos con 5 actores profesionales: dos hombres y tres mujeres con edades comprendidas entre los 20 y 24 años. A estas 20 grabaciones audiovisuales se les extrajo el audio con el fin de constituir un grupo de grabaciones de audio para el test de percepción auditiva.

A continuación se presentan en la figura 1 dos imágenes de muestra extraídas de las grabaciones en las que se pueden apreciar los rasgos faciales distintivos producidos por uno de los actores mientras emitía la palabra *María* de la frase *María se vino desde Caracas solo para verlo* en sus dos modalidades: declarativa e interrogativa.

Figura 1. Expresión facial de uno de los actores durante la emisión de la palabra *María* de la frase *María se vino desde Caracas solo para verlo*. A la izquierda la emisión declarativa y a la derecha la interrogativa.



## 2.2. El test de percepción auditiva: diseño

Este test está constituido por 20 estímulos, correspondientes a los audios extraídos de las 20 grabaciones audiovisuales de enunciados declarativos e interrogativos (cf. tabla 2). Para su elaboración se distribuyeron aleatoriamente los 20 estímulos.

Tabla 2. Estímulos que constituyen el test de percepción auditivo.

ESTÍMULO		
Número	Enunciado	Tipo de frase
1	En Venezuela se habla español	Declarativa
2	Su familia es de Mérida	Declarativa
3	Vas a venir esta tarde	Interrogativa
4	Le robaron la cartera en el cine	Declarativa
5	Tú crees que Rafael es un buen amigo	Declarativa
6	Su hermana ya compró las cosas de la fiesta	Interrogativa
7	Las clases comienzan el sábado	Interrogativa
8	María se vino desde Caracas solo para verlo	Declarativa
9	Enrique estaba solo cuando ocurrió	Interrogativa
10	Los perros destrozaron el jardín anoche	Interrogativa
11	En Venezuela se habla español	Interrogativa
12	Vas a venir esta tarde	Declarativa
13	Su familia es de Mérida	Interrogativa
14	Le robaron la cartera en el cine	Interrogativa
15	Tú crees que Rafael es un buen amigo	Interrogativa
16	Su hermana ya compró las cosas de la fiesta	Declarativa
17	Las clases comienzan el sábado	Declarativa
18	María se vino desde Caracas solo para verlo	Interrogativa
19	Enrique estaba solo cuando ocurrió	Declarativa
20	Los perros destrozaron el jardín anoche	Declarativa

### 2.3. El test de percepción audiovisual: diseño

10 de las 20 grabaciones audiovisuales fueron seleccionadas aleatoriamente y manipuladas con el programa Magic Video Delux: a cada una de ellas se les sustituyó el audio original por el de la misma frase emitida con la modalidad sintáctica contraria. Así, si el enunciado registrado en la grabación era del tipo declarativo, se le sustituyó el audio por el del mismo enunciado emitido de forma interrogativa, obteniendo con ello un estímulo con expresión facial de *declarativa* y audio de *interrogativa*. De esta manera, este test está constituido por 20 estímulos: 10 modificados y 10 no modificados (cf. tabla 3).



Tabla 3. Estímulos que constituyen el tests de percepción audiovisual. En las casillas blancas aparecen los estímulos no modificados y en las casillas sombreadas los estímulos modificados.

ESTÍMULOS			
Número	Frase	Video	Audio
1	En Venezuela se habla español	Declarativa	Interrogativa
2	Su familia es de Mérida	Declarativa	Declarativa
3	Vas a venir esta tarde	Declarativa	Interrogativa
4	Le robaron la cartera en el cine	Declarativa	Interrogativa
5	Tú crees que Rafael es un buen amigo	Declarativa	Interrogativo
6	Su hermana ya compró las cosas de la fiesta	Interrogativa	Interrogativa
7	Las clases comienzan el sábado	Interrogativa	Declarativa
8	María se vino desde Caracas solo para verlo	Declarativa	Declarativa
9	Enrique estaba solo cuando ocurrió	Interrogativa	Interrogativa
10	Los perros destrozaron el jardín anoche	Interrogativa	Interrogativa
11	En Venezuela se habla español	Interrogativa	Interrogativa
12	Vas a venir esta tarde	Declarativa	Declarativa
13	Su familia es de Mérida	Interrogativa	Declarativa
14	Le robaron la cartera en el cine	Interrogativa	Interrogativa
15	Tú crees que Rafael es un buen amigo	Interrogativa	Declarativa
16	Su hermana ya compró las cosas de la fiesta	Declarativa	Declarativa
17	Las clases comienzan el sábado	Declarativa	Interrogativa
18	María se vino desde Caracas solo para verlo	Declarativa	Interrogativa
19	Enrique estaba solo cuando ocurrió	Declarativa	Declarativa
20	Los perros destrozaron el jardín anoche	Interrogativa	Declarativa

#### 2.4. Aplicación de los tests

Los dos tests de percepción fueron aplicados a 200 personas (104 hombres y 96 mujeres), con un promedio de edad de 25,4 años, todos hablantes de español venezolano.

En el test de percepción auditiva, cada uno de los estímulos fue presentado dos veces, dejando 30 segundos de pausa entre ellos. Después de la segunda presentación, los participantes debían seleccionar de entre dos opciones posibles (declarativa o interrogativa) el tipo de frase percibida en cada uno de los estímulos. Para la aplicación se debió usar un computador y unas cornetas a 65dB.

Por su parte, en el test de percepción audiovisual se le pidió a los participantes que seleccionaran de entre dos opciones posibles (declarativa o interrogativa) el tipo de enunciado percibido en cada uno de los estímulos. Las grabaciones audiovisuales fueron proyectadas a los participantes en una pantalla con un video beam y unas cornetas a 65dB. Cada uno de los estímulos fue presentado dos veces, dejando 30 segundos de pausa entre ellos. Al igual que en el test de percepción auditiva, después de la segunda presentación, los participantes debían seleccionar de entre dos opciones posibles (declarativa o interrogativa) el tipo de enunciado percibido en cada uno de los estímulos.

## 2.5. El procesamiento acústico

Cada una de las 20 grabaciones de audio fue procesada con el programa Praat para extraer los valores acústicos de frecuencia fundamental, intensidad y duración correspondientes a los enunciados declarativos e interrogativos con el fin de establecer los valores promedios de estos parámetros acústicos para cada tipo de enunciado. En la tabla 4 se puede apreciar que los enunciados declarativos presentan menor frecuencia fundamental y mayor duración, mientras que lo contrario sucede con los enunciados interrogativos: presentan mayor frecuencia fundamental y menor duración. Los valores de intensidad para ambos tipos de enunciados son casi similares. Estos datos vienen una vez más a confirmar lo que se ha sostenido en otras investigaciones para el español (cf. apartado 1.2.).

Tabla 4. Valores promedios para los enunciados declarativos e interrogativos del español venezolano.

	F0 (Hz)	Duración total (seg)	Intensidad (dB)
Declarativos	198,93	2.04	74,02
Interrogativos	210,21	1.81	73,18

## 3. RESULTADOS Y DISCUSIÓN

### 3.1. El análisis estadístico

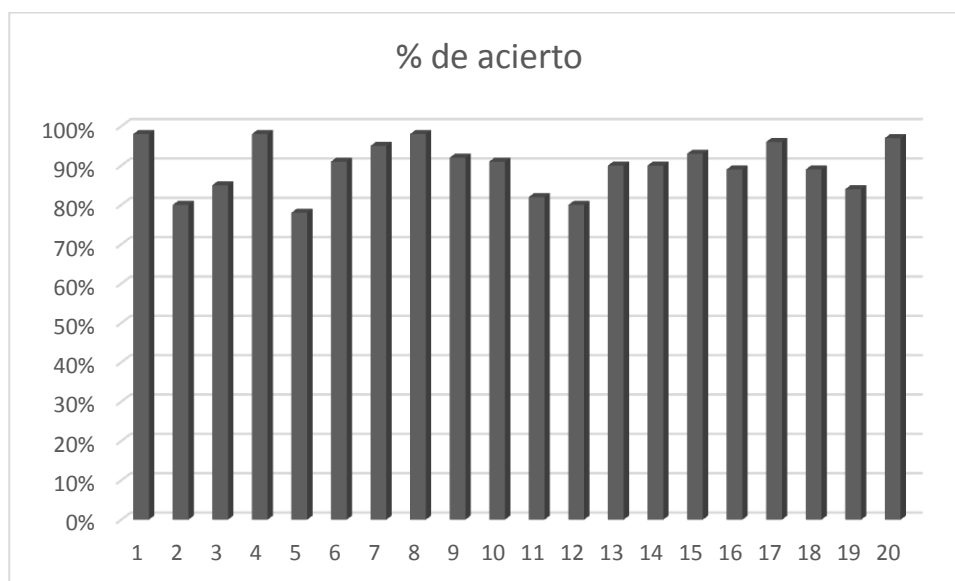
Con el fin obtener el índice de confiabilidad del test audiovisual se aplicó el coeficiente Alfa de Cronbach con el programa SPSS versión 17.0. El resultado de esta prueba arroja un valor de ,939. Esto se traduce en un 93% de confiabilidad de los datos obtenidos del test. Por su parte, el valor arrojado por la misma prueba para el test de percepción auditiva es de ,919. Esto se traduce en un 91% de confiabilidad de los datos obtenidos con este test.

### 3.2. El test de percepción auditivo

Hemos procesado los datos obtenidos del test auditivo en el gráfico 1. De ellos se desprende, en primer lugar, que todos los estímulos, independientemente de si son declarativos o interrogativos, fueron percibidos en altos porcentajes por los 200

participantes del test. El porcentaje de acierto es de 89,8. Los estímulos 1, 4 y 8 han sido los mejores percibidos con un 98% (es decir que 196 participantes del test percibieron correctamente estos estímulos declarativos), mientras que los estímulos 2, 12 y 5 han sido los que han obtenido el menor porcentaje de percepción correcta: 80% para los estímulos 2 y 12 y 78% para el estímulo 5 (esto equivale a decir que 160 participantes percibieron correctamente los estímulos 2 y 12 y solo 156 de los participantes han percibido correctamente el estímulo 5, todos contentivos de enunciados declarativos). Los enunciados declarativos han obtenido el mismo porcentaje de percepción correcta que los interrogativos (89,8%).

Gráfico 1. Resultados generales del test de percepción auditivo. Se indica el porcentaje de acierto para cada estímulo del test de percepción auditivo.

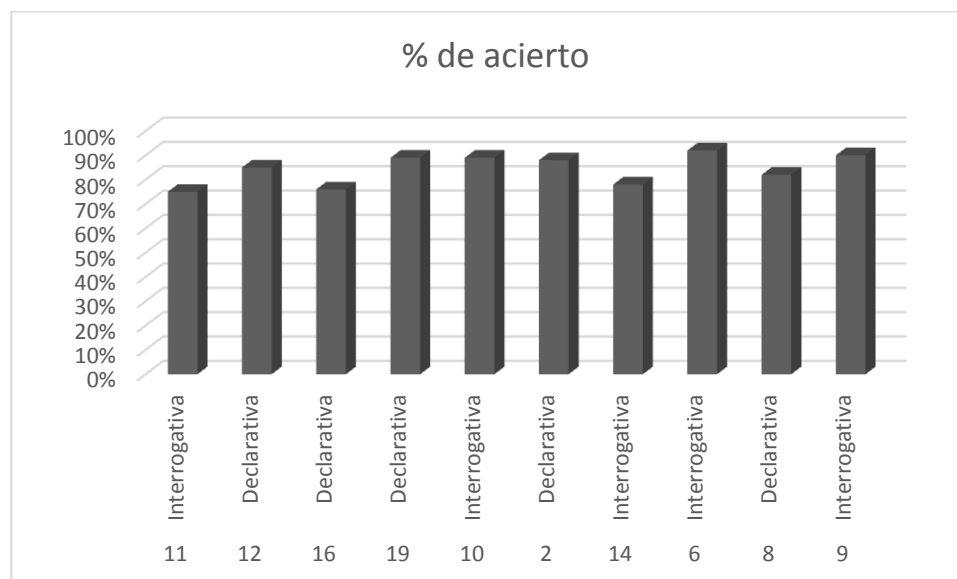


Con estos datos podemos afirmar que los participantes del test han podido percibir en la onda sonora de declarativos e interrogativos la distinción acústica entre estos dos tipos de enunciados, tal como lo señalan los datos acústicos comentados en el apartado 1.2. Los porcentajes de percepción acertada de los estímulos nos permiten decir que los participantes han utilizado los índices acústicos presentes en la señal. En consecuencia, tal como lo afirman Mora *et al.* (2008), la tarea perceptiva se realizó únicamente sobre la base de la señal acústica disponible.

### 3.3. El test de percepción audiovisual

Para la presentación de los resultados de este test discriminaremos entre los obtenidos para los estímulos no-modificados y los modificados, con el fin de determinar si hay alguna influencia de la información visual sobre la percepción de la información auditiva. Los datos obtenidos para los estímulos no-modificados se presentan en el gráfico 2.

Gráfico 2. Datos obtenidos para los estímulos no-modificados del test de percepción audiovisual.

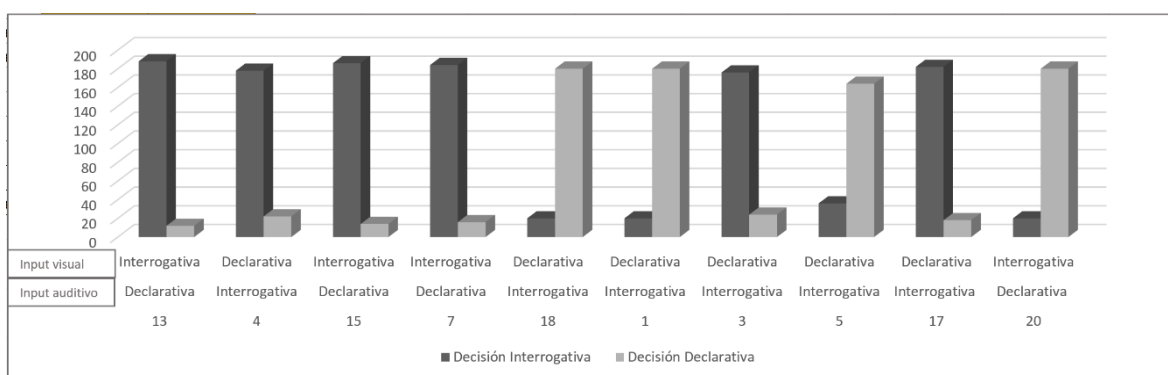


Los porcentajes de percepción correcta del tipo de enunciado en este test audiovisual disminuyen un poco si los comparamos con los datos obtenidos por los mismos estímulos presentados en 3.2., pero en todo caso, siguen siendo porcentajes de acierto bastante altos. Podemos apreciar también en este gráfico que para los estímulos audiovisuales 6 y 9 los participantes acertaron en porcentajes altos al seleccionar como enunciado percibido el interrogativo (92 y 90% respectivamente). Por su parte, los estímulos audiovisuales 11, 16 y 14 obtuvieron los más bajos porcentajes de percepción acertada (75, 76 y 78% respectivamente). El hecho de que estos valores porcentuales obtenidos para los estímulos audiovisuales sean inferiores a los obtenidos para los mismos estímulos auditivos contradeciría el planteamiento formulado por Borrás *et al.* (2013) y Crespo *et al.* (2013) según el cual la precisión de la percepción se mejora aún más cuando se añade información visual.

En el gráfico 3 podremos apreciar los datos obtenidos para los estímulos audiovisuales modificados. A partir de ellos se puede apreciar que para el estímulo 13 (input visual correspondiente al enunciado emitido de manera interrogativa e input sonoro correspondiente al mismo enunciado emitido de manera declarativa), 188 participantes (94%) escogieron como opción el enunciado interrogativo, y 12 participantes (6%), el tipo declarativo como enunciado percibido. Datos similares se han obtenido para el estímulo 15 (input visual interrogativo e input sonoro declarativo): 186 participantes (93%) escogieron el tipo interrogativo y 14 (7%) el tipo declarativo; y para el estímulo 7 (input visual interrogativo e input sonoro declarativo): 184 participantes (92%) escogieron el tipo interrogativo y 16 (8%) el tipo declarativo. Al analizar cada uno de estos casos, vemos que los participantes del test han basado su decisión perceptiva en el input visual y no en el auditivo. Parece que los rasgos que caracterizan la configuración facial de las frases interrogativas (fruncimiento del ceño e inclinación de la cabeza y la no elevación de cejas) tiene un peso importante a la hora de tomar estas decisiones perceptivas.

Esta preferencia por el input visual también puede apreciarse en los datos obtenidos para los estímulos 18, 1 y 5. En el primero de ellos (input visual declarativo e input sonoro interrogativo) pudimos apreciar que la tendencia es a percibirlo como declarativo, 180 (90%) participantes así lo han decidido, mientras que 20 (10%) participantes lo han percibido como interrogativo. Los mismos datos se han obtenido para el estímulo 1 (input visual declarativo e input sonoro interrogativo). 164 (82%) de los participantes de nuestro test han percibido que el estímulo 5 (input visual declarativo e input sonoro interrogativo) también contiene un enunciado declarativo. Los rasgos presentes en la configuración facial que aparece durante la emisión de enunciados declarativos (la elevación de cejas y ausencia de fruncimiento del ceño y de la inclinación de la cabeza) tiene un valor fundamental para los perceptores.

Gráfico 3. Datos obtenidos para los estímulos modificados del test de percepción audiovisual.



Por su parte, en el estímulo 4 de nuestro test (input visual declarativo e input sonoro interrogativo), 178 participantes (89%) escogieron el tipo interrogativo como enunciado percibido en el estímulo, mientras que 22 personas (11%) escogieron como enunciado percibido una declarativo. Este mismo hecho ocurrió con dos estímulos: el estímulo 3 (input visual declarativo e input sonoro interrogativo), en el que 176 (88%) participantes percibieron el estímulo como interrogativo y 24 (12%) lo percibieron como declarativo; y el estímulo 17 (input visual declarativo e input sonoro interrogativo): 182 (91%) personas lo percibieron como interrogativo mientras que 18 (9%) lo percibió como declarativo. De estos datos se desprende que los participantes basaron la percepción de estos estímulos en lo sonoro y no en lo visual. A pesar de que en los videos de cada uno de estos estímulos la configuración facial se correspondía con la de la emisión de un enunciado declarativo (caracterizados por la elevación de cejas, ausencia de fruncimiento del ceño y de la inclinación de la cabeza), la información sonora tuvo mayor peso a la hora de tomar una decisión perceptiva.

De acuerdo con los resultados de nuestro test, los perceptores atienden en primer lugar a la información visual por sobre la información acústica. Cuando la información visual producida durante la emisión del enunciado no es clara para los perceptores, atienden entonces a la información acústica, tal como lo señalan Martínez *et al.* (2012).

Así, de los datos contenidos en el gráfico 3 se desprende que 6 de los estímulos han sido seleccionados por los participantes con base en la información visual y 4 con base en la información acústica.

Pareciera entonces que los enunciados interrogativos, justamente por ser la forma *marcada* en la lengua, suelen percibirse mejor cuando se cuenta con información visual (cf. Higuera, 2013).

#### **4. CONCLUSIONES**

Nuestro objetivo inicial era determinar hasta qué punto los oyentes hacen uso de la información visual para establecer la modalidad sintáctica a la que pertenecen los enunciados declarativos e interrogativos. Luego de la aplicación de los tests afirmamos junto con otros investigadores (cf. Soto, 2000; Borrás *et al.*, 2013; Crespo *et al.*, 2013, por ejemplo) que el acceso a la información visual durante la percepción del habla influye sobre el resultado final del proceso de percepción del habla.

Hemos dicho que en la onda sonora de declarativas e interrogativas se encuentra codificada la distinción acústica entre estos dos tipos de enunciados. Sin embargo, cuando los perceptores cuentan con información visual asociada a la onda sonora de declarativas e interrogativas, esa información puede modificar el resultado del proceso perceptivo contundentemente, sobre todo en los enunciados declarativos. Así, a través de nuestros datos hemos logrado establecer una relación entre el input visual y el auditivo durante la percepción de este tipo de enunciados.

Estos hechos nos llevan a sugerir junto con Collignon *et al.* (2008), Robinson y Sloutsky (2004), Swerts y Krahmer (2005) y Dijkstra *et al.* (2006) la percepción de la información visual es dominante cuando se presentan estímulos incongruentes, pero debemos sugerir también que este dominio no ocurre de manera rígida, sino flexible. Esta afirmación nos lleva a agregar, junto con Martínez *et al.* (2012), que el producto del proceso de percepción audiovisual es una correlación del procesamiento de las informaciones visual y sonora. Los perceptores están cotejando siempre ambos inputs durante el proceso de percepción del habla. Por tanto, confirmamos que la atención entonces recaerá en aquel input que mejor se adapte a las expectativas (contextuales/situacionales, culturales, sociales, anímicas, etc.) del receptor. Siguiendo a Massaro (1987:85), tanto la información auditiva como la visual influyen en la percepción de la prosodia; “la influencia de una modalidad (auditiva o visual) será mayor cuando la información proporcionada por la otra modalidad sea ambigua”.

#### **5. AGRADECIMIENTO**

Esta investigación fue realizada con el financiamiento del Consejo de Desarrollo Científico, Humanístico, Tecnológico y de las Artes (CDCHTA) de la Universidad de Los Andes a través del proyecto H-1438-13-06-B.

## 6. REFERENCIAS BIBLIOGRÁFICAS

Astruc, L.; E. Mora & S. Rew. 2010. Venezuelan Andean Spanish Intonation. En P. Prieto & P. Roseano (eds.). *Transcription of Intonation of the Spanish Language*, 191-226. München: Lincom Europa.

Bertelson, P.; J. Vroomen, B. De Gelder & J. Driver. 2000. The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics* 62: 321-332.

Borrás, J.; C. Kaland, P. Prieto & J. Swerts. 2013. Audiovisual correlates of interrogativity: A comparative analysis of Catalan and Dutch. *Journal of Nonverbal Behavior* 38, 1 : 53-66.

Cavé C. 2003. Illusions auditives. *Dossier Pour la Science* 39 : 60-63.

Cangemi, F. & M. D'Imperio. 2013. Tempo and perception of sentence modality. *Journal of the Association for Laboratory Phonology* 4, 1: 191-219.

Collignon, O.; S. Girard, F. Gosselin, S. Roy, D. Saint-Amour, M. Lassonde & F. Lepore. 2008. Audio-visual integration of emotion expression. *Brain research* 1242: 126-135.

Crespo, V.; C. Kaland, M. Swerts & P. Prieto. 2013. Perceiving incredulity: The role of intonation and facial gestures. *Journal of pragmatics* 47: 1-13.

Dijkstra, C.; E. Kraemer & M. Swerts. 2006. Manipulating uncertainty: The contribution of different audiovisual prosodic cues to the perception of confidence. *Proceedings of the 3rd International Conference on Speech Prosody*. <http://arno.uvt.nl/show.cgi?fid=95687> [12/03/2012].

Dorta, J.; C. Díaz, E. Mora, C. Jorge & N. Rojas. 2013. Comparación de la entonación canario-venezolana. La frecuencia fundamental (F0). En J. Dorta (ed.). *Estudio comparativo preliminar de la entonación de Canarias, Cuba y Venezuela*, 171-216. Tenerife: La página.

Esteve-Gibert, N. & P. Prieto. 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research* 1, 56: 850-864.

Frota, S.; J. Butler & M. Vigário. 2013. Infants' perception of intonation: Is it a statement or a question? *Infancy* 19, 2: 194-213. <http://dx.doi.org/10.1111/infa.12037> [6/10/2015].

Gussenhoven, C. & A. Chen. 2000. Universal and language-specific effects in the perception of question intonation. *International Conference on the Processing of Spoken Language* 6, 2: 1-4.

Higuera, L. 2013. *Influencia de la información visual durante la percepción de frases interrogativas y declarativas del español venezolano*. Memoria de grado. Mérida: Universidad de Los Andes.

Jaekl, P.; A. Alsius A., Pesquita, K. Munhall & S. Soto. 2011. Configural vs motion processing in audiovisual enhancement of speech. Actas de la 34<sup>th</sup> European Conference on Visual Perception 28 August – 1 September 2011. *Perception*, 40-79.

Latinus, M.; F. Joassin, R. Watson, I. Charest, S. Love, P. Mcaleer & P. Belin. 2011. Conjoint and independent neural coding of bimodal face/voice identity investigated with fMRI. Actas de la 34<sup>th</sup> European Conference on Visual Perception 28 August – 1 September 2011. *Perception* 40, 139.

Macdonald I. & H. McGurk. 1978. Visual influences on speech perception process. *Perception & Psychophysics* 24: 253-257.

Maltin, M. & H. Foley. 1996. *Sensación y percepción*. México: Prentice Hall.

Marrero, V. 2001. *Fonética perceptiva*. Madrid: Universidad Nacional de Educación a Distancia.

Martínez, H. 2009. *Evaluación de la percepción y producción del habla*, tesis doctoral. Mérida: Universidad de Los Andes.

Martínez, H; D. Rojas & F. Suárez. 2012. Influencia de la información visual durante la percepción de la prosodia de las emociones actuadas. *Estudios de Fonética Experimental*, XXI: 163-193.

Massaro, D. 1987. *Speech perception by ear and by eye*. Hillsdale, NJ: Erlbaum.

Massaro, D. & M. Cohen. 1983. Integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 9: 753-771.

Méndez, J. 2010. Interacción de los parámetros acústicos duración y frecuencia fundamental en frases declarativas neutras e interrogativas absolutas de los andes venezolanos. *Estudios de Fonética experimental* XIX: 147-164.

Mora, E.; N. Rojas, J. Méndez & H. Martínez. 2008. Declarativas e interrogativas del español venezolano. Percepción de la emisión con y sin contenido léxico. *Language Design* 2: 231-238.

Ouni, S.; M. Cohen, H. Ishak & D. Massaro. 2006. Visual contribution to speech perception: Measuring the intelligibility of animated talking heads. *Journal on Audio, Speech, and Music Processing* 17: 1-12.



Pietrosemoli, L. & E. Mora. 1998. Synchronisation des mouvements des mains et de la ligne de fréquence fondamentale en espagnol parlé. *Actes du Congrès International Orage'98 ORAlité et Gestualité, Communication multimodale et interaction*. 555-559.

Prieto, P.; J. Borràs-Comes, S. Tubau & M. Espinal. 2013. Prosody and gesture constrain the interpretation of double negation. *Lingua* 131: 136-150.

Robinson, C. & V. Sloutsky. 2004. Auditory dominance and its change in the course of development. *Child Development* 75: 1387-1401.

Sekiyama, K.; I. Kanno, S. Miura & Y. Sugita. 2003. Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*,47: 277-287.

Soderstrom, M.; E. Ko & U. Nevzorova. 2011. It's a question? Infants attend differently to yes/no questions and declaratives. *Infant Behavior and Development*, 34, 107-110. <http://dx.doi.org/10.1016/j.infbeh.2010.10.003> [20/05/2015].

Soto, M. 2000. *Influencia de la percepción visual del rostro del hablante en la credibilidad de su voz*, tesis doctoral. Barcelona: Universitat Autònoma de Barcelona.

Srinivasan, R. & D. Massaro. 2003. Perceiving prosody from the face and voice. Distinguishing statements from echoic questions in English. *Language and Speech* 46: 1-22.

Swerts, M. & E. Krahmer. 2005. Audiovisual prosody and feeling of knowing», *Journal of Memory and Language* 53: 81-94.

Tubau, S.; S. González-Fuente, P. Prieto & M. Espinal. 2014. Prosody and gesture in the interpretation of yes-answers to negative yes/no-questions. *The Linguistic Review* 31, 4: 781-808.

Yang, X. & J. Liang. 2012. Declarative and interrogative intonations by brain-damaged speakers of Uygur and Mandarin Chinese. En Q. Ma, H. Ding & D. Hirst (eds.). *Proceedings of the Sixth International Conference on Speech Prosody*,701-704. Shanghai.

Yeni-Komshian, G. 1999. Percepción del habla. En J. Berko Gleason & N. Bernstein (eds.). *Psicolingüística*. Madrid: Mc Graw Hill.

Yuan, J.; C. Shih & G. Kochanski. 2002. Comparison of declarative and interrogative intonation in Chinese. En B. Bel & I. Marlien (eds.). *Proceedings of the First International Conference on Speech Prosody*, 711-714. Aix-en-Provence.