

**INFLUENCIA DE LA INFORMACIÓN VISUAL  
DURANTE LA PERCEPCIÓN DE LA PROSODIA  
DE LAS EMOCIONES ACTUADAS**

**INFLUENCE OF VISUAL INFORMATION  
DURING THE PERCEPTION OF PROSODY  
OF ACTUATED EMOTIONS**

HERNÁN MARTÍNEZ

*Universidad de los Andes (Mérida, Venezuela)*

*hjmart@gmail.com*

DARCY ROJAS

*Universidad de los Andes (Mérida, Venezuela)*

*darcyjra@gmail.com*

FÉLIZ SUÁREZ

*Universidad de los Andes (Mérida, Venezuela)*

*felixfelicis3@gmail.com*

*Artículo recibido el día: 11/06/2012*

*Artículo aceptado definitivamente el día: 10/09/2011*

*Estudios de Fonética Experimental, ISSN 1575-5533, XXI, 2012, pp. 163-193*

---

**RESUMEN**

La integración audiovisual durante la percepción del habla ha sido analizada en varias investigaciones. Se ha afirmado que el acceso a la información visual durante la percepción del habla proporciona inteligibilidad a la percepción del mensaje sonoro. Desde esta perspectiva, en esta investigación se trata de determinar, a través de análisis perceptivos, la influencia que ejerce el acceso a la información visual durante el proceso de percepción de la prosodia de las emociones. Para lograr este objetivo hemos diseñado un corpus conformado por muestras audiovisuales de habla controlada obtenidas de registros con actores profesionales. Posteriormente, parte de esas muestras fueron manipuladas. Diseñamos un experimento de percepción en el que 152 sujetos participaron escuchando y viendo las muestras audiovisuales para luego determinar la emoción que percibían. A través del análisis de los resultados pudimos determinar que los oyentes evalúan las informaciones de la fuente auditiva y de la visual, la disponibilidad de información visual modifica la percepción de la información acústica. Así concluimos que el producto del proceso de percepción audiovisual es una correlación del procesamiento de las informaciones visual y sonora.

Palabras clave: *emociones, información visual, percepción del habla.*

**ABSTRACT**

The fact of the audiovisual integration during the speech perception has been stated in several researches. It has been argued that access to visual information during speech perception provides intelligibility in the perception of sound message process. From this perspective, this research is to determinate, through perceptive analysis, the influence that the access to the visual information performs during the process of perception of the prosody of emotions. To reach this aim, we have built a corpus of controlled audiovisual speech samples. We got these samples from professional actors' registers. Later, parts of those samples were manipulated. We designed an experiment for perception with 152 participants. They had to listen and look at the audiovisual samples in the perception test so we could determinate the emotion that they perceived and select it among the options we present to them. Through the analysis of the result we could determinate that the listerners evaluate the information from the visual and auditory source. The availability of visual information changes the perception of acoustic information. To conclude, the product of visual perception process is a correlation of visual information processing and sound.

Keywords: *emotions, visual information, speech perception.*

## 1. LA PERCEPCIÓN AUDIOVISUAL DEL HABLA

La percepción, en tanto que función que nos permite, por medio de los sentidos, interpretar, asignar significado y organizar las experiencias generadas por algún tipo de estímulo, se caracteriza por ser un fenómeno amplio e integral en el que influyen nuestras necesidades, valores y deseos. Cada uno de nosotros codifica y clasifica los datos que nos llegan del exterior, reduciéndolos a categorías de las que dispondremos para comprender luego el entorno. Estas clasificaciones y codificaciones son procesos intermediarios entre los estímulos y la conducta (Bruner 1984). Desde esta perspectiva, se reconoce, en efecto, que la percepción es un proceso compuesto de sensaciones (objetivas) y de conocimientos previos (subjetivos) que se nos aparece como simple; esto es, todo lo percibido parece estar dado simultáneamente y en igualdad de condiciones, formando una unidad. Específicamente, la percepción del habla la definimos, siguiendo a Martínez (2009:35) como *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*.

Desde el punto de vista lingüístico, varios han sido los estudios sobre percepción del habla. En ellos este proceso se ha vislumbrado bastante complejo, así lo indican, entre otros, autores como Maltin y Foley (1996), Yeni-Komshian (1999), Marrero (2001) y Martínez (2009). Los trabajos de percepción del habla 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 (aún 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.

Parte de las investigaciones sobre percepción del habla se ha centrado exclusivamente en la señal acústica. En parte, la concentración en la percepción auditiva está justificada por el hecho de que la señal acústica es rica en información sobre el lenguaje hablado. Sin embargo, la señal visual puede tener gran impacto en la percepción de la señal auditiva bajo condiciones degradantes. Así,

*The relative importance of the visual signal increases as the auditory channel is degraded through noise, distortion, filtering, hearing loss, and potentially through*

---

---

*unfamiliarity with a particular talker, stimulus set, or listening condition. When information in the visual channel is in disagreement with the information in the auditory channel, the visual channel may change or even override the percept of the auditory channel alone (Cleary y Pisoni 1998:35).*

La psicología experimental ha acumulado a lo largo de varios años sobradas evidencias empíricas que le han permitido afirmar que la disponibilidad de información visual sobre un hablante modifica la percepción de la información acústica de su 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 más citada sobre este aspecto es el efecto o ilusión McGurk (MacDonald y McGurk 1978). Con ese nombre se conoce a la demostración del fenómeno audiovisual en la percepción. Es la primera evidencia experimental de la integración audiovisual del habla. Los resultados del trabajo de estos investigadores demuestran, por ejemplo, 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 (se podría afirmar que las señales visuales y auditivas del experimento son dispares; que no pertenecen a un único evento del habla, por lo que generan un efecto subjetivo: una ilusión de ventriloquismo. La integración 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*). A pesar de esto, en el mismo experimento MacDonald y McGurk mostraron que el efecto de la integración audiovisual, aunque menos pronunciado, también se obtenía con la presentación visual y sonora de sílabas concordantes.

En realidad, MacDonald y McGurk pretendían asignar un papel a la visión en la percepción del habla. Los investigadores manifestaban que las teorías descansaban sobre el paradigma de que el proceso era unimodal y básicamente sonoro. Los escasos estudios que asignaban un papel a la visión en el habla partían desde la óptica de que:

1. Su tarea era alternativa a la audición (en sujetos con discapacidades, por ejemplo); o
2. su tarea era complementaria (aportaba inteligibilidad en situaciones de degradación sonora).

Según los resultados obtenidos por MacDonald y McGurk, en un nivel de procesamiento superior, existe una relación interactiva entre oído y visión por la que se integra y sintetiza la información disponible en ambos canales. Investigaciones posteriores comprobaron que la integración audiovisual del habla es un fenómeno sólido e inmune a una variedad de situaciones experimentales. Por ejemplo, este fenómeno es sólido 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; o si la presentación de la información acústica y visual no se corresponde con un hablante del mismo género. 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 (cf. Soto 2000). Todas estas pruebas son, en sí mismas, un ejemplo de que la presentación del efecto McGurk avivó el interés de un grupo de investigadores interesados en el procesamiento perceptivo del habla audiovisual.

Así, desde el punto de vista metodológico, Massaro y Cohen (1983) justifican el uso de la ilusión y la combinación factorial de las fuentes auditivas, visuales y audiovisuales como la manera apropiada de estudiar la interacción entre los sistemas perceptivos auditivo y visual. En 1983, Massaro y Cohen reportaron los resultados de un experimento mediante el cual se proponían describir la manera en que las informaciones auditivas y visuales se integraban en el habla. La primera observación de los investigadores fue 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<sup>1</sup>. En el mismo estudio, Massaro y Cohen afirman que los movimientos articulatorios visibles durante la percepción del habla aportan inteligibilidad a la decodificación del mensaje sonoro en situaciones de degradación acústica<sup>2</sup>. 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,

---

<sup>1</sup> Los autores comprobaron que el receptor obtenía informaciones diferentes de cada canal en la identificación de las señales del habla bimodal. A los sujetos el canal visual les servía para conocer el lugar en el que el hablante realizaba la articulación, mientras que el sonoro les enseñaba la forma en que el hablante articulaba.

<sup>2</sup> Por ejemplo, se ha podido demostrar que el acceso a la información visual es de mucha utilidad durante la percepción del habla en ambientes ruidosos (Cf. Massaro 1987; Assmann y Summerfield 2004).

---

en muchos casos, los segmentos fonéticos tienen equivalentes en los movimientos de la boca. Concluyen los autores 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, Vroomen, De Gelder y Driver (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.

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.

## **2. LAS EMOCIONES: PROSODIA Y EXPRESIÓN FACIAL**

El uso de las emociones en el habla tiene como finalidad, entre otras cosas, servir de herramienta para enriquecer de manera paralela al mensaje, haciendo que un enunciado pueda ser emitido de diferentes formas y ser entendido según la cantidad de veces que el emisor haya usado un tipo de emoción, sin verse alterado el contenido semántico, mas no pragmático, del mensaje. Según Scherer (2005) las emociones son definidas como episodios de cambios coordinados en varios componentes (incluyendo al menos activación neuropsicológica, expresión motriz y sentimientos subjetivos pero posiblemente también tendencias a la acción y procesos cognitivos) en respuesta a acciones externas.

### **2.1. La prosodia**

Scherer (1982: 568) indica que la expresión oral de las emociones viene cargada de *un carácter analógico-vocal conectado al mecanismo biológico del individuo*. Según el autor, *en las características vocales, el hablante, nos brinda información extra inconscientemente sobre sus emociones*; así manifiesta que las expresiones de

tristeza en una persona pueden identificarse porque el habla es lenta y por producir una mala articulación. Las emociones, entonces, son susceptibles de diferenciarse entre sí por una serie de elementos lingüísticos. Múltiples investigaciones han demostrado que cada tipo de emoción posee características acústicas específicas que, como hablantes competentes de una determinada variedad lingüística, compartimos. Motivo por el cual podemos, como emisores, comunicar un determinado estado emocional y, como receptores, reconocerlo (Cf. Scherer 2003:231).

Los diferentes estudios llevados a cabo siguiendo el método de emociones simuladas (Jiménez 1987; Johnstone y Scherer 2000; Scherer, Johnstone y Klasmeyer 2003; Viola y Madureira 2007; Fricke 2008; Martínez y Rojas 2011) revelan datos que parecen apuntar a que algunos tipos de emoción cuentan con patrones acústicos que les permiten diferenciarse entre sí. Entre los parámetros analizados en el estudio de las emociones se encuentran: el valor promedio de la frecuencia fundamental (F0), el rango de la frecuencia fundamental, la variabilidad melódica, la intensidad y la duración. Según Blondet (2006:14-15), la frecuencia fundamental (F0), la melodía, el ritmo y la intensidad de la voz cambian con el estado emocional, evidenciando de manera inequívoca lo que el individuo siente en ese momento.

Desde esta perspectiva, Martínez y Rojas (2011), quienes han analizado acústicamente con Praat la rabia, la alegría y la tristeza en el español hablado en Mérida (Venezuela) en 15 grabaciones interpretadas por 4 actores profesionales, sostienen que cada una de las emociones presenta características acústicas distintas (Cf. Tabla 1).

	F0 (Hz)	INTENSIDAD (dB)	DURACIÓN TOTAL (S)
RABIA	240,32	77,51	1,42
ALEGRÍA	202,83	74,81	1,08
TRISTEZA	183,77	61,68	1,17
NEUTRO	176,84	71,77	1,12

Tabla 1. *Parámetros acústicos de las declarativas neutras y de cada una de las emociones analizadas en español venezolano.*

Con base en sus resultados, Martínez y Rojas (2011) afirman que el dominio de la frecuencia fundamental (F0) tiene una importancia particular para la codificación de la rabia, igualmente la intensidad tiene un papel fundamental. La rabia parece estar caracterizada por un incremento de la F0 y de la intensidad. Como veremos en la tabla 2, en la rabia también se puede apreciar un incremento de la variabilidad y del rango de la F0. Igualmente, afirman Martínez y Rojas que:

1. La alegría presenta un incremento de la frecuencia fundamental y de la intensidad.
2. En la tristeza hay un descenso de la intensidad y de la F0.
3. La rabia y la alegría se alejan en mayor medida de la media de la F0 de las declarativas neutras, mientras que la tristeza se acerca más a la media de estas.

	FRECUENCIA FUNDAMENTAL (F0)			DURACIÓN		
	VARIABIL. (%)	TONO PROMEDIO (Hz)	DESV. TÍPICA (Hz)	VARIABIL. (%)	DURACIÓN PROMEDIO (S)	DESV. TÍPICA (S)
RABIA	10,19	240,32	24,49	10,19	1,42	0,0611
ALEGRÍA	10,3	202,83	20,91	7,41	1,08	0,0801
TRISTEZA	5,27	183,77	9,7	5,96	1,17	0,0698
NEUTRO	32,03	176,84	56,65	2,4	1,12	0,0269

Tabla 2. Variabilidad melódica y temporal de las emociones y las declarativas neutras.

Así mismo, sostienen Martínez y Rojas (2011) que la velocidad del habla parece ser fundamental en la diferenciación acústica, y por tanto perceptiva, de las emociones. De este modo, la velocidad de habla asciende en función del tipo de emoción: la rabia se acerca a los rangos de la velocidad de habla lenta, mientras

que la tristeza se acerca más a los rangos de la velocidad de habla normal y la alegría, por su parte, coincide con los valores de la velocidad de habla rápida<sup>3</sup>.

En el mismo estudio, Martínez y Rojas (2011), luego de realizar un test de percepción a 225 personas, sostienen que la emoción mejor percibida es la tristeza, seguida por la rabia y la alegría<sup>4</sup> (Cf. figura 1). En la investigación se sostiene que los oyentes tienden a confundir, aunque en bajos porcentajes la rabia con la alegría y viceversa (28% y 33%, respectivamente). La misma no discriminación se aprecia entre la tristeza y la alegría (9%)<sup>5</sup>. Según los autores, estas no discriminaciones podrían obedecer, en ambos casos, a la similitud entre los valores de  $f_0$  e intensidad que presentan ambos grupos de emociones. Es probable que para algunos de los oyentes *no haya quedado clara la variabilidad melódica (que, acústicamente, nos permite diferenciar a la tristeza del resto de las emociones analizadas) y la variabilidad temporal (que nos permite distinguir, acústicamente, entre la rabia y la alegría)* (Martínez y Rojas 2011:70).

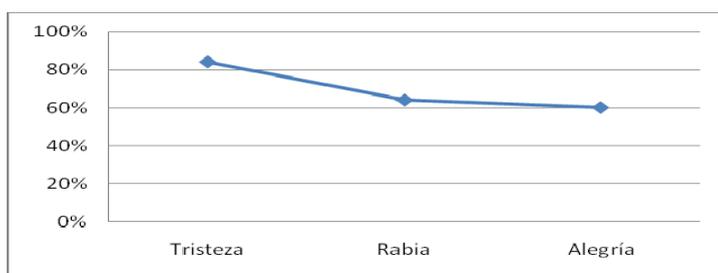


Figura 1. Porcentajes de identificación de las emociones por parte de los oyentes.

<sup>3</sup> Los valores de velocidad lenta, normal y rápida a los que se hace referencia han sido establecidos por Blondet (2006).

<sup>4</sup> Afirman Jiménez y Mallo (1988) en su estudio que la alegría es la emoción más difícil de percibir. Scherer (2003), indica que la rabia es la emoción que mayor grado de exactitud obtiene en una gran cantidad de estudios que ha revisado, seguida por la tristeza.

<sup>5</sup> Según Scherer (2003), los oyentes parecen tener problemas para identificar de forma inequívoca las emociones solamente a través de la voz. Según este investigador, la tristeza y la ira son generalmente los mejores perceptibles, seguidas por el miedo. En varias investigaciones se ha determinado que contar con el soporte visual de la emoción ayuda en gran medida a la percepción inequívoca de las emociones.

---

## 2.2. La expresión facial

Además de la voz, los movimientos del rostro humano son también vehículo de las emociones y han sido objeto de estudio de diversas disciplinas, entre ellas la historia natural y la etología. Ya en el siglo XIX, Charles Darwin (1872. Citado por Chóliz 1995; Krauss, Chen y Chawla 1996) argumentaba a favor de la universalidad de las expresiones faciales de las emociones. Para Darwin, los movimientos del rostro que evocaban el estado de ánimo de las personas tenían una expresión común en todas las culturas, debido a que ellos son vestigios de adaptación de la especie humana. Por ejemplo, la expresión de desagrado puede manifestarse literal o metafóricamente como un acto de repulsión del gusto y el olfato, frunciendo ligeramente la nariz para cerrar las fosas nasales; también, con el miedo, la sangre se retira del rostro (lo que explica la palidez y la sensación de *quedarse frío*) y fluye a otras partes del cuerpo, como las piernas, favoreciendo así la huida (Conde López 2005).

Ekman y Oster (1981) reseñan también diferentes trabajos e investigaciones del siglo XX en los cuales se comprueba que las expresiones faciales pueden considerarse universales, dado que la manifestación y la interpretación de algún gesto facial no son diferentes entre diversas culturas (por ejemplo, americanas y asiáticas, literatas y no literatas), y lo que influye en la manifestación o en la interpretación diversa es el aprendizaje cultural sobre comportamientos que son innatos.

Por su parte, Ray Birdwhistell (1979) asegura que los seres humanos somos capaces de detectar los movimientos faciales y, con ello, las emociones de las personas. Por ejemplo, suponemos que una persona está triste, feliz o enojada por una serie de movimientos de la cara que nos permiten inferir tales estados de ánimo. Gracias a esta suposición, Birdwhistell explicaba que cada expresión del rostro podía analizarse en unidades distintivas más pequeñas (los quinemas, que serían los equivalentes a los fonemas en una lengua natural). Esta propuesta no tuvo repercusiones positivas dentro de la lingüística, pero es base fundamental en los estudios de psicología y de ingeniería computacional (debido, sobre todo, a la recreación de gestos faciales en programas informáticos), y, hoy por hoy, la hipótesis de una base innata en el ser humano en la producción de gestos faciales para expresar emociones no es tema de discusión.

En cuanto a los movimientos que, en conjunto, expresan una emoción específica, Chóliz (1995) hace referencia a varios de ellos (el asco, el miedo, la sorpresa). Solo

daremos cuenta aquí de las emociones alegría, tristeza y rabia, que son las que analizamos en este trabajo y parte de las emociones que Chóliz también analiza:

*1. Alegría:*

- Elevación de las mejillas.
- Comisura labial retraída y elevada.
- Arrugas en la piel debajo del párpado inferior.

*2. Tristeza:*

- Ángulos inferiores de los ojos hacia abajo.
- Piel de las cejas en forma de triángulo.
- Descenso de las comisuras de los labios, que incluso pueden estar temblorosos.

*3. Rabia:*

- Cejas bajas, contraídas y en disposición oblicua.
- Párpado inferior tenso.
- Labios tensos, o abiertos, y en ademán de gritar
- Mirada prominente.

Todas estas marcas, en conjunto, pueden llevarnos a suponer cuál es el estado de ánimo de la persona que estemos observando.

Para nuestro estudio, sin embargo, debemos partir de 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 de estado de ánimo específico. Para ello, consideraremos cuatro variables: movimiento del ceño, movimiento de las cejas, movimiento de la comisura de los labios, y tensión de los labios. Esta matriz de rasgos no pretende ser rigurosa, entre otras cosas porque no estamos considerando otras emociones (como la sorpresa, el miedo o el desagrado), pero para entender los estados de ánimo que estamos estudiando, son suficientes los cuatro rasgos señalados. Así, tendríamos en la tabla 3 las cuatro emociones analizadas en esta investigación y sus respectivos rasgos:

	FRUNCIMIENTO DEL CENEO	LEVANTAMIENTO DEL EXTREMO EXTERIOR DE LAS CEJAS	LEVANTAMIENTO DEL EXTREMO INTERNO DE LAS CEJAS	ARRUGA DEL PÁRPADO INFERIOR	DESCENSO DE LA COMISURA ORBICULAR	LEVANTAMIENTO DE LA COMISURA DE LOS LABIOS	TENSIÓN DE LOS LABIOS
ALEGRÍA	(-)	(-)	(+)	(+)	(-)	(+)	(-)
RABIA	(+)	(+)	(-)	(+)	(-)	(-)	(+)
TRISTEZA	(+)	(-)	(+)	(-)	(+)	(-)	(-)
NEUTRO	(-)	(-)	(-)	(-)	(-)	(-)	(-)

Tabla 3. Rasgos gestuales para cada una de las emociones. Los símbolos + y - señalan la presencia o la ausencia del rasgo, respectivamente.

Como vemos en la tabla 3, las tres primeras emociones presentan cada una un conjunto de rasgos distintos, mientras que el estado emocional neutro no presenta ninguno de esos rasgos. Con ello, podemos diferenciar, desde el punto de vista de la articulación gestual, cada una de las emociones.

### 3. LA PERCEPCIÓN AUDIOVISUAL DE LA PROSODIA DE LAS EMOCIONES

La información sobre los estados emocionales puede ser comunicada a través de una variedad de comportamientos no verbales, tales como las expresiones faciales (Rothman y Nowicki 2004). Aunque la percepción audiovisual de las emociones es un tema poco explorado, varias han sido las investigaciones que se han preocupado por determinar el rol de la expresión facial durante la percepción de la prosodia de las emociones. Massaro y Egan (1996) y Hietanen, Leppänen, Illi y Surakka (2004) han mostrado a través de varios experimentos que cuando a los sujetos se les presentan combinadas, por ejemplo, una expresión prosódica de *rabia* en una modalidad (auditiva) y una expresión de *alegría* en otra modalidad (visual) cambian las respuestas a *alegría* y viceversa. De Gelder y Vroomen (2000) obtuvieron resultados similares con las siguientes combinaciones: *alegría-tristeza*, *rabia-tristeza* y *alegría-miedo*. Aplicando un experimento similar, Fagel (2006) obtuvo resultados poco claros para la percepción de los estímulos en los que se combinaban dos emociones, mientras que los estímulos audiovisuales congruentes en los que no se combinaban emociones fueron mejor percibidos por los participantes en la investigación. Sostiene Fagel (2006) que durante la percepción audiovisual las señales son ponderadas en las modalidades auditiva y visual y luego integradas en la percepción de una emoción.

Collignon *et al.* (2008) señalan que en el proceso de percepción audiovisual de emociones, cuando se presentan estímulos incongruentes en los que se intercambian el audio de una emoción por el video de otra, se puede observar un claro dominio de la información visual; esto también ha sido afirmado por Robinson y Sloutsky (2004), Swerts y Kraimer (2005) y Dijkstra *et al.* (2006). Sin embargo, los investigadores encontraron que cuando el estímulo visual no tenía buena calidad expresiva, los participantes clasificaban los estímulos con base en la expresión auditiva. Sugieren Collignon *et al.* (2008) que el dominio de la percepción visual no ocurre de manera rígida, sino flexible, depende mucho de la situación en la que se dé el proceso de percepción. Sobre esto, Massaro (1987) en el *Modelo de Lógica Difusa de Percepción*, afirma que 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 sea ambigua.

---

En este mismo sentido, Borràs-Comes y Prieto (2011), sostienen que la integración audiovisual es necesaria para que el procesamiento de la percepción de la prosodia sea rápido y preciso. Al analizar a través de dos experimentos el rol de los gestos visuales (congruentes e incongruentes) durante la interpretación de tonos de declarativas e interrogativas en catalán, Borràs-Comes y Prieto han encontrado una relevancia importante y constante de la señal visual en las decisiones de los oyentes. Sin embargo, no descartan la importancia de la señal acústica en el proceso de interpretación de las declarativas e interrogativas en catalán cuando el input visual es incongruente.

Barbosa, Vatikiotis-Bateson y Yehia (2002) al estudiar la relación entre la señal acústica del habla y los movimientos de los músculos faciales afirman que el habla tiene una naturaleza multimodal en la que los componentes acústicos y visuales trabajan juntos en el proceso de transmisión de la información. Es posible entonces, sugerir que la información fonética visual es una consecuencia de la configuración del tracto bucal para generar la señal acústica del habla, indicando que los eventos visuales y acústicos compartirían la misma fuente neuromotora.

Desde esta perspectiva, en esta investigación experimental se trata de determinar, a través de análisis perceptivos y estadísticos, la influencia que ejerce el acceso a la información visual durante el proceso de percepción de la prosodia de las emociones, concretamente queremos saber:

1. ¿A cuál de los dos sentidos involucrados en el proceso de percepción (auditivo y visual) atienden los oyentes?
2. ¿Cuál es la contribución de cada uno de los sentidos involucrados (visual y auditivo) en la percepción de la prosodia de las emociones? Partimos de la hipótesis de que existe un claro dominio de la información visual durante la percepción del habla.

## **4. METODOLOGÍA**

### **4.1. Corpus**

Con la intención de determinar si la expresión facial influye en la percepción de la expresión hablada de las emociones, se trabajó con una parte de un corpus audiovisual recogido por integrantes y colaboradores del Grupo de Investigaciones en Ciencias Fonéticas de la Universidad de Los Andes. Este corpus está constituido

por una serie de grabaciones audiovisuales en las que se recogen emociones simuladas por 5 actores profesionales: dos hombres y tres mujeres con edades comprendidas entre los 20 y 24 años.

Para el desarrollo de la investigación se seleccionaron del corpus 8 grabaciones audiovisuales en las que los cinco actores producían (a petición del responsable de la grabación) el enunciado *prepara una torta* de manera alegre, triste, rabiosa y neutra<sup>6</sup>. Una vez registradas, esas 8 grabaciones fueron manipuladas con el programa Magic Video Delux: a cada una de ellas se les sustituyó el audio original por el de otra emoción. Así, si la emoción registrada en la grabación era *alegría*, se le sustituyó el audio por el de *rabia*, obteniendo con ello un estímulo con expresión facial de *alegría* y audio de *rabia*. De esta manera, la base de datos estuvo constituida, entonces, por 16 grabaciones de enunciados declarativos, 8 manipulados y 8 originales (Cf. tabla 4).

Nº DEL ESTÍMULO	EXPRESIÓN FACIAL	AUDIO
1	Alegría	Rabia
2	Rabia	Rabia
3	Neutro	Neutro
4	Tristeza	Tristeza
5	Rabia	Alegría
6	Tristeza	Tristeza
7	Alegría	Tristeza
8	Tristeza	Neutro
9	Alegría	Alegría
10	Neutro	Neutro
11	Rabia	Rabia

<sup>6</sup> Los parámetros acústicos asociados a cada una de las emociones analizadas han sido descritos y expuestos en el apartado 2 según un estudio previo (Cf. Martínez y Rojas 2011).

12	Rabia	Neutro
13	Alegría	Rabia
14	Tristeza	Rabia
15	Alegría	Alegría
16	Alegría	Neutro

Tabla 4. Estímulos que constituyen el corpus de la investigación. En las casillas blancas aparecen los estímulos originales y en las casillas sombreadas los estímulos manipulados.

A continuación se presentan en la figura 2 cuatro imágenes de muestra extraídas de las 8 grabaciones del corpus en las que se pueden apreciar las expresiones faciales (y sus rasgos distintivos) de los actores mientras emitían la primera sílaba de la frase *prepara una torta* con las distintas emociones.



Figura 2. Expresiones faciales de los actores mientras emiten la primera sílaba de la frase *Prepara una torta*. De izquierda a derecha y de arriba a abajo están representadas las siguientes emociones: neutro, alegría, tristeza y rabia.

#### **4.2. Test de percepción**

Un test de percepción audiovisual se diseñó especialmente para la investigación. El test está constituido por 16 ítems que se corresponden con los 16 estímulos que constituyen el corpus y la duración aproximada de aplicación fue de 10 minutos. Para su realización se distribuyeron aleatoriamente los 16 estímulos. Cada estímulo fue presentado dos veces, dejando 30 segundos de pausa entre cada uno de ellos. Después de la segunda presentación, las personas debían seleccionar de entre cuatro opciones posibles (alegría, tristeza, rabia o neutro) la emoción percibida en cada una de las grabaciones o estímulos.

Este test se aplicó a varios grupos de personas de manera secuencial en el Laboratorio de Fonética de la Universidad de Los Andes. Para la proyección de las grabaciones a los participantes se utilizó una pantalla, un video beam y unas cornetas a 65dB.

Una vez aplicado el test se procedió a calcular el porcentaje de acierto para cada uno de los 16 estímulos del test.

#### **4.3. Participantes**

El test de percepción lo realizaron 152 personas (74 hombres y 78 mujeres), con un promedio de edad de 22,7 años, todos hablantes de español venezolano y estudiantes de la Facultad de Humanidades y Educación de la Universidad de Los Andes.

#### **4.4. Test estadístico**

Con el fin obtener el índice de confiabilidad del test se aplicó el coeficiente Alfa de Cronbach con el programa SPSS versión 17.0. Además, con este mismo programa se hizo una prueba de correlación de ítems para ver cuáles de los 16 estímulos contenidos en el test de percepción pudieran ser rechazados y con ello mejorar la confiabilidad del mismo.

## 5. RESULTADOS

A partir de los datos arrojados por la aplicación del test podemos observar, en primer lugar, que todas las emociones, indistintamente de si son o no modificadas, fueron percibidas en altos porcentajes por los participantes del test. Sin embargo, para el análisis, se discriminó entre los resultados obtenidos para los estímulos modificados y los no-modificados, con el fin de determinar si había alguna influencia de la información visual sobre la percepción de la información auditiva. Los datos obtenidos para los estímulos manipulados se presentan en el gráfico de la figura 3.

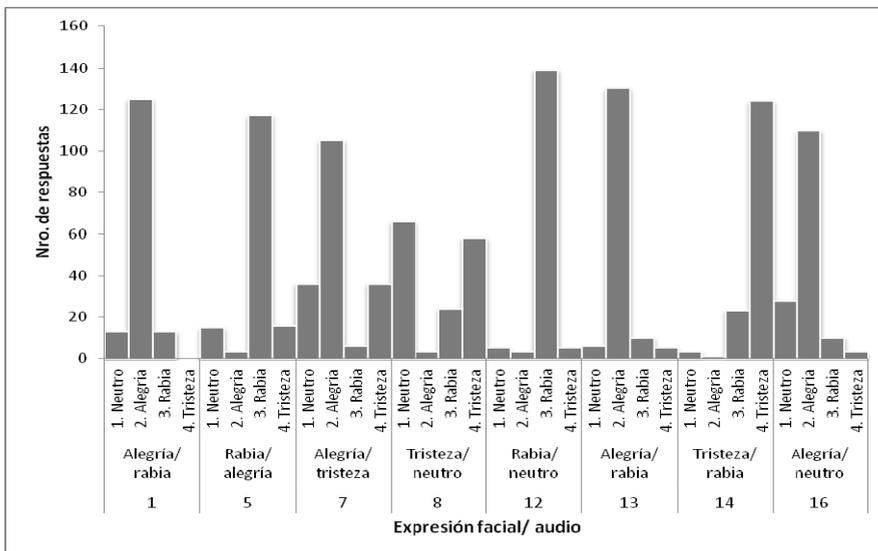


Figura 3. Resultados del test de percepción de emociones. Los ítems 1, 5, 7, 8, 12, 13, 14 y 16 corresponden a los estímulos manipulados.

A partir de estos datos se puede apreciar que para el estímulo 1 (input visual de alegría e input sonoro de rabia), 125 participantes (82%) escogieron como emoción la alegría, y 13 participantes (9%), la rabia como emoción percibida. En el estímulo 5 de nuestro test (input visual de rabia e input sonoro de alegría), 117 participantes (77%) escogieron la rabia como emoción percibida en el estímulo, mientras que

solo 3 personas (2%) escogieron como emoción la alegría. Por su parte, en el estímulo 7 (input visual de alegría e input sonoro de tristeza), el margen de diferencia fue menor: 105 personas (57%) escogieron la emoción de la alegría y 36 participantes (20%), la de la tristeza. Otro 20% de los participantes optó por la emoción neutra. Este margen de diferencia menor entre los resultados obtenidos para la alegría (el input visual) y las emociones tristes y neutras se debe probablemente a la poca enfatización, por parte del actor del video, en los rasgos faciales para la alegría (precisamente la elevación de las comisuras labiales). Sin embargo, la diferencia de respuestas para este estímulo sigue siendo considerable.

Resultados diferentes conseguimos en el estímulo 8 (input visual de tristeza e input sonoro de voz neutra). En esta parte del test, 66 participantes (equivalente a 44%) escogieron la opción del input sonoro y 58 personas (38%), el input visual de la tristeza. Esto quizás se deba, también, a la poca expresividad del actor del video para mostrar tristeza (en la expresión facial, todos los rasgos fueron negativos, lo que equivale a la expresión neutra). Los resultados en el estímulo 12 (input visual de rabia e input sonoro neutro) muestran que 139 personas (91%) escogieron la rabia como emoción percibida, mientras que solo 3% de la muestra escogió la emoción neutra. En el estímulo 13 (input visual de alegría e input sonoro de rabia) 130 personas (86%) escogieron la alegría como emoción percibida; 10 personas (7%) escogieron la rabia. En el estímulo 14 (input visual de tristeza e input sonoro de rabia), 124 personas (82%) escogieron la tristeza, mientras que 23 personas (15%), optaron por la rabia. Por último, en el estímulo 16 (input visual de alegría e input sonoro neutro) encontramos que 110 participantes (77%) escogieron la alegría como emoción percibida; y 28 personas (19%), la emoción neutra.

Estos resultados nos indican que la mayoría de los participantes selecciona la rabia con mayor facilidad cuando su input es visual. Para los estímulos 5 y 12, cuyo estímulo visual era la rabia, la selección de ésta como emoción percibida fue considerable (88% y 91% respectivamente). Esto puede deberse, en parte, al fruncimiento del ceño, pero sobre todo al levantamiento del lado exterior de las cejas y a la tensión de los labios, características que no encontramos en las otras emociones estudiadas.

Con respecto a la alegría, podemos decir que los participantes también escogían mayormente esta emoción cuando su input es visual, aunque la diferencia de resultados con respecto a las otras emociones, fue menor, sobre todo en los estímulos 7 y 16, en cuyos videos no se enfatiza el movimiento que, quizás, diferencia esta emoción de cualquier otra: el levantamiento de la comisura de los labios. Esto hizo que, en estos estímulos, los participantes escogieran, en una proporción considerable, la opción de emoción neutra, cuyos rasgos son de valor negativo.

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 de la emoción no es claro para los perceptores, atienden entonces a la información acústica. Esto se puede apreciar, por ejemplo, en los resultados obtenidos para el estímulo 14, cuyos inputs visual y acústico son, respectivamente, la tristeza y la rabia. En este caso el 15% de los participantes del test escogieron la rabia como emoción percibida.

Sin embargo, cuando los valores de producción son similares, tanto en lo visual como en lo acústico (como la tristeza y lo neutro), los oyentes pueden confundir la emoción que perciben. Uno de los estímulos estudiados, el estímulo 8, presenta un input visual de tristeza y un input acústico neutro, y los resultados para este estímulo son similares para ambas emociones, se dividen entre la información visual (38% para la tristeza) y la información auditiva (43% para el neutro).

En la figura 4 se presentan los datos obtenidos para los estímulos originales, es decir, estímulos en los que tanto el audio como la expresión facial se corresponden.

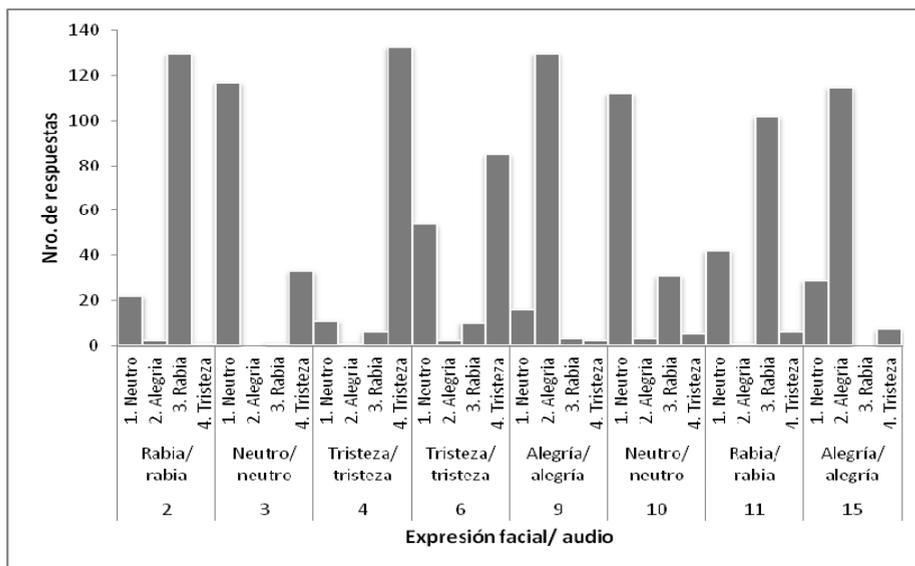


Figura 4. Resultados del test de percepción de emociones. Los ítems 2, 3, 4, 6, 9, 10, 11 y 15 corresponden a los estímulos originales.

Los datos contenidos en este gráfico nos permiten apreciar que para los estímulos 1 y 11 los participantes en el test acertaron al seleccionar la rabia como la emoción percibida (86 y 67% respectivamente). Vale la pena resaltar el hecho de que para ambos estímulos hubo quienes optaron por seleccionar la emoción neutro en porcentajes considerables. En el caso del estímulo 11 estamos de acuerdo en afirmar que el 28% de los participantes optó por la emoción neutro porque las expresiones faciales del actor no son de la rabia (caracterizada por presentar fruncimiento del ceño, levantamiento del extremo exterior de las cejas, arrugas del párpado inferior, tensión de los labios; y por no presentar levantamiento del extremo interno de las cejas, descenso de la comisura orbicular y levantamiento de la comisura de los labios).

Con respecto a los estímulos 4 y 6, en los que la emoción correcta era la tristeza, encontramos que es muy bien percibida en el primer estímulo (88%) mientras que en el estímulo 6 se confunde con la emoción neutro (56% para la tristeza y 36 % para neutro). Esto quizás se deba, también, a la poca expresividad del actor del video para mostrar tristeza (en la expresión facial, todos los rasgos fueron negativos, lo que equivale a la expresión neutra). Sin embargo, parece ser que los oyentes en este caso se dejaron llevar por la prosodia y no por la expresividad facial.

La emoción alegría fue muy bien percibida en los estímulos 9 y 15 (86% y 76% respectivamente). Estos estímulos originales son los que mejor porcentajes de percepción presentan. Este hecho parece confirmar la hipótesis de que el acceso a la información visual influye en la percepción de la prosodia de las emociones, pues ya afirmaban Jiménez y Mallo (1988) y Martínez y Rojas (2011), que la alegría es la emoción menos percibida en los test en los que solo se presentan estímulos acústicos.

Altos porcentajes de acierto también se obtuvieron para los estímulos 3 y 10 en los que la emoción correcta era la neutra (77% y 74% respectivamente). Sin embargo, se pudo observar que existe un poco de confusión con la tristeza, a pesar de que tanto el input acústico y el visual no se han modificado. Con base en esta última afirmación, sostenemos que los oyentes se dividen entre estas dos emociones por las características tanto acústicas como expresivas que ambas comparten (Cf. datos de Martínez y Rojas 2011 citados en el apartado 2).

Afirmaremos entonces lo que Scherer (2003) ya ha planteado: contar con el soporte visual de la emoción ayuda en mucho a la percepción inequívoca de estas. No sucede lo mismo que cuando solo se cuenta la prosodia. Esta afirmación la

corroboran los datos estadísticos: según el test t de student (Cf. tabla 5), las medias de las respuestas de los estímulos manipulados y originales no son diferentes significativamente ( $p > .35$ ). Con base en este dato, sostenemos que tanto en estímulos manipulados como en los originales la información visual es importante para la percepción de las emociones.

**Prueba de muestras independientes**

	Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
							95% Intervalo de conf. para la dif.		
	F	Sig.	t	gl	Sig. (bilateral)	Dif.de medias	Error típ. de la diferencia	Inferior	Superior
RESP. Se han asumido varianzas iguales	,847	,357	6,122	2439	,000	,28540	,04662	,19399	,37681
No se han asumido varianzas iguales			6,123	2438,995	,000	,28540	,04661	,19400	,37681

Tabla 5. Prueba t de Student para observar las diferencias obtenidas al comparar los estímulos manipulados y los originales.

Por otra parte, para obtener el índice de confiabilidad del test se aplicó el coeficiente Alfa de Cronbach (Cf. tabla 6) con el programa SPSS v. 17, el resulta-

do obtenido del análisis de confiabilidad arroja un valor de ,536. Lo cual se traduce en un 53% de confiabilidad del test, tomando como base los 16 estímulos que lo componen.

ALFA DE CRONBACH	N DE ELEMENTOS
,536	16

Tabla 6. *Índice de confiabilidad del test.*

La correlación de cada ítem con el total, reseñado en la columna 4 de la tabla 7, evidencia por su parte, que de los 16 estímulos, 4 se encuentran por debajo de ,30; y por tanto, deberían ser desechados para mejorar la confiabilidad del test. No obstante, como se puede observar en la columna 5, que refleja el alfa de Cronbach, solo la eliminación de tres de ellos permitiría un ascenso considerable del nivel de confiabilidad de la prueba.

	ESTADÍSTICOS TOTAL-ELEMENTO			
	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
ESTÍMULO1	37,7566	25,351	,447	,503
ESTÍMULO2	37,0592	22,056	,725	,434
ESTÍMULO3	38,0921	36,495	-,708	,717
ESTÍMULO4	36,0263	20,423	,868	,391
ESTÍMULO5	36,8684	21,969	,733	,432
ESTÍMULO6	36,9211	19,875	,441	,444
ESTÍMULO7	37,2434	22,954	,432	,472
ESTÍMULO8	37,2763	20,347	,412	,455
ESTÍMULO9	37,8158	24,747	,593	,489

ESTÍMULO10	38,2434	32,503	-,568	,650
ESTÍMULO11	37,2895	22,763	,412	,473
ESTÍMULO12	35,9211	22,351	,803	,437
ESTÍMULO13	37,6645	27,695	-,106	,553
ESTÍMULO14	36,2632	19,811	,647	,401
ESTÍMULO15	37,8553	24,204	,472	,485
ESTÍMULO16	38,0526	33,547	-,545	,683

Tabla 7. *Correlación de los ítems del test con el total.*

Una vez eliminado estos 3 estímulos, la confiabilidad del test ascendió a ,888; es decir a un 88 %, tal como se muestra en la tabla 8.

ESTADÍSTICOS DE FIABILIDAD	
ALFA DE CRONBACH	N DE ELEMENTOS
,888	13

Tabla 8. *Índice de confiabilidad del test una vez eliminados los estímulos 3, 10 y 16.*

Vale la pena destacar que este conjunto de tres estímulos, susceptibles de ser descartados del test, se corresponden con emociones con características acústicas y expresivas semejantes a las de otra emoción. En los estímulos 3 y 10 la emoción contenida es neutra (cuyas características acústicas y expresivas son semejantes con las de la tristeza). Lo mismo sucede en el estímulo 16 donde la emoción contenida en la alegría y se confunde con la neutra.

Finalmente, es necesario agregar que después de analizar los resultados del test, se comprobó a través de la prueba t de Student que las diferencias entre las medias obtenidas al considerar la variable sexo no son significativas: ,616 (Cf. tabla 9).

	Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
								95% Intervalo de conf. para la dif.	
	F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	Inferior	Superior
RESP. Se han asumido varianzas iguales	,252	,616	1,188	2438	,235	,05582	,04699	-,03633	,14796
No se han asumido varianzas iguales			1,188	2428,317	,235	,05582	,04698	-,03630	,14794

Tabla 9. Diferencias obtenidas al considerar la variable sexo en los resultados del test de percepción.

## 6. CONCLUSIONES

Nuestro objetivo inicial era determinar a cuál de los dos sentidos involucrados en el proceso de percepción (auditivo y visual) atienden los oyentes. Después del análisis de los resultados podemos sostener que los oyentes 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

---

oyente. Sin embargo, nuestros datos nos hacen pensar que los oyentes atienden mejor al input visual que al auditivo durante la percepción audiovisual.

Con bases en los resultados presentados, podemos responder también a la pregunta número 2 que nos hiciéramos anteriormente: ¿cuál es la contribución de cada uno de los sentidos involucrados (visual y auditivo) en el reconocimiento de la prosodia de las emociones? En el proceso de percepción audiovisual, el input visual tiene un gran peso en la percepción de la prosodia de las emociones. Los datos muestran que éste 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 hablantes seleccionan la información visual en un porcentaje superior al 70%; esto se evidencia en los estímulos en los que los pares alegría/rabia, rabia/alegría, alegría/tristeza, rabia/neutro, alegría/rabia, tristeza/rabia y alegría/neutro aparecen. Sin embargo, también podemos afirmar que cuando el estímulo visual no cuenta con suficiente calidad expresiva, los informantes optan por la información auditiva-prosódica. 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) que el dominio de la percepción visual es indudable, pero debemos afirmar también que este dominio no ocurre de manera rígida, sino flexible. Esta afirmación nos lleva a agregar que el producto del proceso de percepción audiovisual es una correlación del procesamiento de las informaciones visual y sonora. Siguiendo a Massaro (1987), tanto la información auditiva como la visual influyen en la percepción de la prosodia; el *Modelo de Lógica Difusa de Percepción* nos permite dar cuenta de que la influencia de una modalidad (auditiva o visual) será mayor cuando la información proporcionada por la otra modalidad sea ambigua.

Es necesario precisar en este punto que no pudimos observar a través de los datos obtenidos si el efecto McGurk está presente en la percepción de la prosodia de las emociones. Esto quizá se deba al hecho de que los participantes del test tenían solo cuatro opciones para elegir una, quedando por fuera la posibilidad de que ellos pudieran informar sobre otra emoción percibida.

*AGRADECIMIENTO: Esta investigación ha podido llevarse a cabo gracias al financiamiento del Consejo de Desarrollo Científico, Humanístico, Tecnológico y de las Artes (CDCHTA) de la Universidad de Los Andes a través del proyecto código H-1275-09-06-B.*

---

## 7. REFERENCIAS BIBLIOGRÁFICAS

- ASSMANN, P. y Q. SUMMERFIELD. (2004): «The perception of speech under adverse conditions», en S. Greenberg y W. Ainsworth (eds.): *Speech processing in the auditory system*, Nueva York, Springer Verlag, pp. 231–308.
- BARBOSA, A.; E. VATIKIOTIS-BATESON y H. YEHIA (2002): «Modeling the relation between speech acoustics and 3D facial motion», *Technical Report of the Institute of Electronics, Information and Communication Engineers*, 102 (735), pp. 13-18.
- BERTELSON, P.; J. VROOMEN; B. DE GELDER y J. DRIVER (2000): «The ventriloquist effect does not depend on the direction of deliberate visual attention», *Perception and Psychophysics*, 62, pp. 321-332.
- BIRDWHISTELL, R. (1979): *El lenguaje de la expresión corporal*, Barcelona, Gustavo Gili SA.
- BLONDET, M. (2006): *Variaciones de la velocidad de habla en español: patrones fonéticos y estrategias fonológicas. Un estudio desde la producción*, tesis doctoral, Universidad de Los Andes (Mérida-Venezuela).
- BORRÀS-COMES, J. y P. PRIETO. (2011): «‘Seeing tunes.’ The role of visual gestures in tune interpretation», *Laboratory Phonology*, 2 (2), pp. 355-380.
- BRUNER, J. (1984): *Acción, pensamiento y lenguaje*, Madrid, Alianza.
- CHÓLIZ, M. (1995): «La expresión de las emociones en la obra de Darwin», en F. Tortosa, M. Civera y C. Catalayud (comps.): *Prácticas de Historia de Psicología*, Valencia, Promolibro.
- COLLIGNON, O.; S. GIRARD; F. GOSSELIN; S. ROY; D. SAINT-AMOUR; M. LASSONDE y F. LEPURE. (2008): «Audio-visual integration of emotion expression», *Brain research*, 1242, pp. 126-135.
- CONDE LÓPEZ, R. (2005): *Mapeo facial de emociones sintéticas*. [http://tierra.aslab.upm.es/documents/PFC/PFC\\_RPConde.pdf](http://tierra.aslab.upm.es/documents/PFC/PFC_RPConde.pdf). [10/10/11].
- CLEARY, M. y D. PISONI. (1998): «Speech perception and spoken word recognition: Research and theory», *Research on spoken language processing*, 22, pp. 56-78.

- 
- DE GELDER, B. y J. Vroomen. (2000): «The perception of emotions by ear and by eye», *Cognition & Emotion*, 14, pp. 289-311.
- DIJKSTRA, C.; E. KRAHMER y 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*, Desden (Alemania).  
<http://arno.uvt.nl/show.cgi?fid=95687> [12/03/2012].
- EKMAN, P. y H. OSTER. (1981): «Expresión facial de las emociones», *Estudios de Psicología*, 7, pp. 116-144.
- FAGEL, S. (2006): «Emotional McGurk effect», *Proceedings of the 3rd International Conference on Speech Prosody*, Desden (Alemania).  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.138.3943&rep=rep1&type=pdf>. [14/10/11].  
[http://sprogis.isle.illinois.edu/sp2006/contents/papers/SPS2-6\\_0006.pdf](http://sprogis.isle.illinois.edu/sp2006/contents/papers/SPS2-6_0006.pdf) [12/03/2012]
- FRICKE, A. (2008): «Relating emotional content to speech rate in Brazilian Portuguese», en P. Barbosa, S. Madureira y C. Reis (eds.): *Proceedings of 4<sup>th</sup> the Speech Prosody 2008 Conference*, Campinas, Editora RG/CNPq, pp. 217-222.
- HIETANEN, J.; J. LEPPÄNEN; M. ILLI y V. SURAKKA. (2004): «Evidence for the integration of audiovisual emotional information at the perceptual level of processing», *European Journal of Cognitive Psychology*, 16, pp. 769-790.
- JIMÉNEZ, F. (1987): *Marcadores emocionales en la conducta vocal*, Madrid, Universidad Autónoma de Madrid.
- JIMÉNEZ, F. y M. MALLO. (1988): «El reconocimiento de emociones a través de la voz», *Estudios de psicología*, 33-34, pp. 31-52.
- JOHNSTONE, T. y K. SCHERER. (2000): «Vocal communication of emotion», en M. Lewis y J. Haviland (eds.): *Handbook of emotions*, Nueva York, Guilford Press, pp. 220-235.
- KRAUSS R.; Y. CHEN y P. CHAWLA. (1996): «Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us?», en M.
-

- 
- Zanna (ed.): *Advances in experimental social psychology*, San Diego, CA, Academic Press, pp. 389-450.
- MALTIN, M. y 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, Universidad de Los Andes (Mérida, Venezuela).
- MARTÍNEZ, H. y D. ROJAS. (2011): «Prosodia y emociones: datos acústicos, velocidad de habla y percepción de un corpus actuado», *Lengua y Habla*, 15, pp. 59-72.  
<http://erevistas.saber.ula.ve/index.php/lenguayhabla/article/viewFile/3356/3257> [15/11/2011].
- MASSARO, D. (1987): *Speech perception by ear and by eye*, Hillsdale, NJ, Erlbaum.
- MASSARO D. y M. COHEN (1983): «Integration of visual and auditory information in speech perception», *Journal of Experimental Psychology: Human Perception and Performance*, 9, pp. 753-771.
- MASSARO, D. y P. EGAN. (1996): «Perceiving affect from the voice and the face», *Psychonomic Bulletin and Review*, 3, pp. 215-221.
- MACDONALD, I. y H. MCGURK. (1978): «Visual influences on speech perception process», *Perception & Psychophysics*, 24, pp. 253-257.
- ROBINSON, C. y V. SLOUTSKY. (2004): «Auditory dominance and its change in the course of development», *Child Development*, 75, 5, pp. 1387-1401.
- ROTHMAN, A. y S. NOWICKI. (2004): «A measure of the ability to identify emotion in children's tone of voice», *Journal of Nonverbal Behavior*, 28, pp. 67-92.
- SCHERER, K. (1982): «Emotion as a process: function, origin and regulation», *Social Science Information*, 21, pp. 550-570.
- SCHERER, K. (2003): «Vocal communication of emotion: A review of research paradigms», *Speech Communication*, 40, pp. 227-256.

- SCHERER, K.; T. JOHNSTONE y G. KLASMEYER. (2003): «Vocal expression of emotion», en R. Davidson, K. Scherer y H. Goldsmith (eds.): *Handbook of the affective sciences*, Nueva York y Oxford, Oxford University Press, pp. 433-456.
- SCHERER, K. (2005): «What are emotions? And how can they be measured?», *Social Science Information*, 44, 4, pp. 695-729.
- SOTO, M. (2000): *Influencia de la percepción visual del rostro del hablante en la credibilidad de su voz*, tesis doctoral, Universitat Autònoma de Barcelona.
- SWERTS, M. y E. KRAHMER. (2005): «Audiovisual prosody and feeling of knowing», *Journal of Memory and Language*, 53(1), pp. 81-94.
- VIOLA, I y S. MADUREIRA. (2007): «Voice quality and speech expressiveness», *Proceedings of the Paralinguistic Speech07-between models and data*, Saarbrücken, DFKI Publication, pp. 65-70.
- YENI-KOMSHIAN, G. (1999): «Percepción del habla», en J. Berko Gleason y N. Bernstein (eds): *Psicolingüística*, Madrid, Mc Graw Hill.