

Prosodia y emociones: datos acústicos, velocidad de habla y percepción de un corpus actuado

Hernán Martínez

Darcy Rojas

Laboratorio de Fonética

Maestría en Lingüística

Universidad de Los Andes. Mérida, Venezuela



Resumen

El estado emocional del hablante suele ser transmitido fielmente a través de la prosodia. Los cambios que se producen en todos los estados emocionales son detectados y cuantificados en el producto acústico final, ya sea a través de diferentes herramientas de análisis acústico o a través de las valoraciones que hacen los oyentes en los tests de percepción. Los objetivos de esta investigación son: (1) revisar y analizar sistemáticamente los correlatos acústicos asociados a emociones actuadas (alegría, tristeza y rabia) y, (2) establecer cuál es la capacidad perceptiva de los oyentes para identificar esas emociones. El corpus utilizado para lograr estos objetivos está constituido por 20 grabaciones de la enunciado *Prepara una torta* interpretadas por 5 actores profesionales. Los actores simulaban cada una de las emociones indicadas durante la emisión del enunciado. Estas grabaciones fueron tratadas con el programa Praat para obtener los correlatos acústicos asociados a cada una de las emociones. Posteriormente, se hizo un test de percepción en el que se le pidió a los informantes que identificaran cada una de las emociones. Los datos obtenidos del análisis acústico muestran que hay diferencias consistentes entre las características acústicas para cada una de los enunciados grabados. A través del test de percepción, se pudo observar que las emociones, en más del 70% de los casos, son identificadas claramente. Sin embargo, se observa la existencia de casos en los que los oyentes fallan en la identificación. Las razones de este hecho se discuten en el texto.

Palabras clave: emociones, prosodia, análisis perceptivo, análisis acústico, velocidad del habla, actuación.

Abstract

Emotions are usually transmitted through prosody. Emotional changes are detected and quantified by various speech tools for acoustic analysis or by the evaluation made by hearers through perception tests. This research study has the following objectives: (1) to review and analyze utterances related with simulated emotions (joy, sadness, and

anger), and (2) to determine the perceptive ability of hearers to identify these emotions. The corpus is based on fifteen recordings of the expression *Prepara una torta* (Bake a cake) played by four professional actors, who simulated each emotion through the same expression. The recordings were then analyzed through Praat software; the acoustic utterances related to emotions were obtained. A perception test was made and the informants were asked to identify each one of the emotions they heard. Data obtained from acoustic analysis showed regular differences among acoustic characteristics for each utterance recorded. Through the perception test, more than 55 percent of emotional utterances were clearly identified. However, some utterances remained unidentified by the hearers. The reasons for this situation are discussed in the paper.

Key words: emotions, prosody, perceptive analysis, acoustic analysis, speech rate, actuation.

1. INTRODUCCIÓN

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.

El estudio de las emociones, y especialmente el estudio de las emociones desde el punto de vista lingüístico, es un tema que ha originado múltiples controversias causadas, bien por la dificultad de distinguir el concepto de las emociones del de otros estados de ánimo que experimenta el ser humano, bien por la dificultad de hallar un método idóneo que permita estudiar este fenómeno desde la producción y la percepción del lenguaje.

Entre las dificultades mencionadas anteriormente la mayor parece ser que, como lo afirma Scherer (2003:242), en esta área parece no existir un consenso que permita dar una definición del fenómeno emocional y de las características que diferencian los distintos tipos de emociones existentes. El problema comienza, como afirma el autor, con la distinción de las emociones de los demás tipos de estados afectivos del hablante, tales como el humor, las actitudes e incluso rasgos afectivos relacionados con la personalidad. 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.

En estudios posteriores de Scherer (1982:558), se indica que la expresión oral de las emociones viene cargado 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.

Debido a los inconvenientes que presenta el estudio de las emociones, Johnstone and Scherer (2000) sugieren algunas características que, en su opinión, permiten distinguirlas de otras clases de estados afectivos. En este sentido, afirman los autores, se

aconseja distinguir los diferentes estados afectivos por la intensidad relativa y por la duración, por la medida en que estos estados buscan provocar una reacción y una valoración, por la rapidez de los cambios de estado y el grado en el cual el estado afecta el comportamiento del individuo.

Las emociones, así clasificadas, se consideran un grupo especial entre las clases de estados afectivos, que se define por su mayor intensidad, su corta duración, por la existencia de un propósito (provocar reacciones y valoraciones determinadas en el interlocutor) y por su gran influencia en el cambio de comportamiento del hablante.

Otra de las premisas que apoya el estudio de las emociones propuesto anteriormente es que estas son susceptibles de diferenciarse entre sí por una serie de elementos tanto gestuales como lingüísticos. Uno de los aspectos que más se ha discutido en este contexto lingüístico se centra en si el estado emocional de un hablante puede ser transmitido a través de la voz. Para determinar la veracidad de esta hipótesis múltiples investigaciones se han propuesto demostrar 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).

Con el fin de comprobar las características acústicas de cada emoción se han utilizado distintos métodos que permiten categorizar el estudio de este fenómeno en tres tipos: los estudios que parten de la expresión espontánea de emociones (utilizando para ello grabaciones de reportajes televisivos o de sesiones terapéuticas), los que parten de expresiones emocionales inducidas a través de diferentes sustancias psicotrópicas, y los estudios que parten de expresiones emocionales simuladas por actores profesionales. Este último método parece tener cierta ventaja y aceptabilidad, pues permite superar algunas desventajas que presentan los métodos anteriores, tales como la inadecuada calidad de la señal que frecuentemente tienen los estudios espontáneos, y que imposibilita el análisis acústico, o la manifestación demasiado sutil de estados emocionales, que se logra con las emociones inducidas, y que puede imposibilitar el reconocimiento de un determinado estímulo por parte de un receptor.

A pesar de estas aparentes ventajas, los estudios que parten de expresiones emocionales simuladas parecen tener también ciertas desventajas, debido a que existe la posibilidad de que los actores puedan enfatizar demasiado algunas características acústicas y obviar otras características probablemente usadas en la expresión natural de las emociones. No obstante, estas dificultades parecen poder subsanarse, al menos en parte, si se realizan estudios complejos que tengan en consideración la valoración de receptores que verifiquen la verosimilitud de cada emoción y las características acústicas más relevantes para su reconocimiento.

Se entiende con el razonamiento expuesto el por qué algunos de los modelos utilizados para la identificación de las características acústicas de las emociones parten de un estudio amplio que considera tanto la producción (análisis computacional de los parámetros acústicos de la señal) como percepción del fenómeno emocional (análisis de parámetros obtenidos a través de pruebas de percepción) (Cf. Scherer, 2003; Muñoz y Jiménez 1990). Pues, además de que este análisis dual permite abordar el estudio de las emociones como un fenómeno funcional e interactivo en el que están presentes emisor y

receptor, también permite estudiar este fenómeno, tan susceptible a la intervención de múltiples variables, de manera más o menos objetiva.

Los diferentes estudios llevados a cabo siguiendo el método de emociones simuladas (Jiménez, 1987; Johnstone and Scherer, 2000; Scherer et al., 2003; Viola y Madureira, 2007; Fricke, 2008) 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, el rango de la frecuencia fundamental, la variabilidad melódica, la intensidad y la duración. Según Blondet (2006:14-15), la frecuencia fundamental, la melodía, el ritmo y la intensidad de la voz cambian con el estado emocional, evidenciando de manera inequívoca lo que ese individuo siente en ese momento.

Otras investigaciones han mostrado cómo las emociones tienen un fuerte efecto sobre la velocidad de habla. Estas investigaciones se han realizado en diferentes lenguas, con distintos métodos, objetivos y propósitos, estudios como los de Banse y Scherer (1996) para el inglés; Burkhardt (2001 citado en Trouvain, 2004), para el alemán y Gimson Ramsaram (1989, citado en Bouzon, 2004:62), para el francés, arrojaron resultados muy similares ya que un incremento en la tasa de velocidad de habla caracterizaba la rabia, el miedo, la ira, e incluso la felicidad, por el contrario, el aburrimiento, la tristeza, el pesar y el disgusto estuvieron caracterizados por una desaceleración de la velocidad de habla.

Para otros investigadores, sin embargo, ha sido evidente que en la caracterización acústica de las emociones deben ser tomadas en consideración, además de las variables acústicas anteriormente descritas, diferencias lingüístico-culturales. La importancia de estas diferencias radica en que estas pueden aportar información relevante, debido a que aparentemente existe una variación de los patrones acústicos que caracterizan cada emoción dependiendo de la cultura. Al menos así lo registran Muñoz y Jiménez (1990) quienes en su exposición de datos de emociones señalan que mientras que en casi todos los estudios anglosajones, y otros pocos en lenguas europeas, emociones como la ira se caracterizan por un alto nivel de la frecuencia fundamental, extenso rango de la frecuencia fundamental, gran variabilidad de la frecuencia fundamental, voz intensa y breve duración; en estudios como los de Jiménez (1987) aplicados al español revelan un nivel de la frecuencia fundamental de acuerdo con la media y una mayor duración respecto a los valores medios de las declarativas neutras.

También en la codificación de otras emociones como la tristeza se muestran patrones distintos dependiendo del idioma, pues mientras que en los estudios realizados para el inglés y para algunas lenguas europeas, se le atribuyen características como: un bajo nivel de la frecuencia fundamental, menor rango de la frecuencia fundamental, menor variabilidad de la frecuencia fundamental, voz menos intensa y mayor duración; en el estudio del español peninsular se le atribuyen características como: nivel de frecuencia fundamental de acuerdo con la media y duración mayor. De la misma manera, la codificación del miedo muestra diferencias significativas: un alto nivel de la frecuencia fundamental, extenso rango de la frecuencia fundamental, gran variabilidad de la frecuencia fundamental, voz intensa y breve duración, para el inglés; y bajo nivel de frecuencia fundamental y duración mayor, para el español peninsular (Cf. Tabla 1).

Tabla 1. Síntesis de los parámetros acústicos de las emociones que, según Muñoz y Jiménez (1990), han sido tratados en español o en otras lenguas.

PARÁMETROS	EMOCIÓN					
	IRA		TRISTEZA		MIEDO	
	ESTUDIOS EN OTRAS LENGUAS	ESTUDIOS EN ESPAÑOL	ESTUDIOS EN OTRAS LENGUAS	ESTUDIOS EN ESPAÑOL	ESTUDIOS EN OTRAS LENGUAS	ESTUDIOS EN ESPAÑOL
F0	MAYOR	MEDIO	MENOR	MEDIO	MAYOR	MENOR
RANGO	MAYOR	- ¹	MENOR	-	MAYOR	-
VARIABILIDAD	MAYOR	-	MENOR	-	MAYOR	-
INTENSIDAD	MAYOR	-	MENOR	-	MAYOR	-
DURACIÓN	MENOR	MAYOR	MAYOR	MAYOR	MENOR	MAYOR

Como se puede observar en la Tabla 1, los datos presentados por Muñoz y Jiménez (1990), registran los parámetros acústicos utilizados más frecuentemente en el estudio de las emociones y los comparan con algunos de los parámetros estudiados por Jiménez (1987). Para establecer la comparación los autores presentan los datos según su posición en una escala. Lo cual, si bien no permite establecer distinciones entre la tristeza y el miedo, en español, permite determinar si existen diferencias entre las emociones y los valores medios de cada parámetro, así como también comprobar las diferencias acústicas que existen entre emociones como la ira y la tristeza en otras lenguas o entre emociones como la tristeza y el miedo en español.

La falta de distinción entre algunas emociones registradas en la investigación de Muñoz y Jiménez (1990), si bien en un primer momento pudiera atribuirse a la ausencia de una descripción numérica de las distintas emociones, se presenta también en estudios descritos numéricamente como el de Beller *et al.* (2006). Este estudio, realizado para el francés, indica que emociones como la alegría y la ira muestran similitudes en el comportamiento de la frecuencia fundamental (250 Hz aprox.), pero se diferencian en parámetros temporales como la velocidad de habla, la cual resulta más rápida para la ira y menos rápida para la alegría.

Los hallazgos encontrados por Beller *et al.* (2006), están lejos de refutar la hipótesis que defiende la existencia de parámetros acústicos característicos de cada emoción. No obstante, confirman la necesidad de tomar en cuenta en el estudio acústico de las emociones patrones temporales como la velocidad de habla pues, en el caso específico de algunas emociones, este parámetro puede permitir una descripción más acertada de las características acústicas.

Particularmente, sobre el análisis fonético de las emociones en el español venezolano parecen no existir antecedentes directos que permitan comparar los patrones temporales e incluso entonativos de los diferentes tipos de emociones. Esta situación, aunada a la cantidad de variables fonéticas, metodológicas, dialectales y lingüísticas en general necesarias de ser tomadas en cuenta para la caracterización de las emociones, nos lleva a delimitar el análisis de las emociones a una serie de variables iniciales que nos permitan aportar una primera descripción del fenómeno en el español de la región.

¹ Las casillas vacías indican los datos no explícitos en la investigación de Muñoz y Jiménez (1990)

Optamos, por lo tanto, por realizar un primer estudio acústico y temporal que, partiendo de la producción, nos permita confirmar o refutar la importancia de los elementos acústicos y temporales en la clasificación de las emociones, y de esta manera orientar la confirmación de los resultados que nos permitan determinar cuáles de los parámetros que caracterizan cada emoción son relevantes en el proceso de percepción del habla.

2. METODOLOGÍA

2.1. El corpus

Con la intención de determinar si los parámetros acústicos y la velocidad de habla funcionan como elementos caracterizadores de los distintos tipos de 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. Ese corpus está constituido por una serie de grabaciones audiovisuales en las que se recogen emociones simuladas por actores profesionales.

Para el desarrollo de la investigación se seleccionaron del corpus 20 grabaciones en las que se producía el enunciado *prepara una torta* de manera alegre, triste, rabiosa y neutra. La muestra final estuvo constituida, entonces, por 20 grabaciones de enunciados declarativos y ciento veinte sílabas. Estos enunciados fueron producidos por cinco hablantes: dos hombres y tres mujeres. Cada uno de los cinco hablantes produjo el enunciado con cada una de las emociones indicadas.

2.2. El análisis acústico

Cada una de las grabaciones audiovisuales fue almacenada en formato .WAV. Posteriormente, a través del Praat, y utilizando un espectrograma de banda ancha, se segmentó cada una de las noventa sílabas contenidas en los 15 enunciados, incluidas las pausas encontradas, y se realizó el cálculo de la intensidad, la frecuencia fundamental y la duración promedio de cada una de los enunciados, además de la duración y frecuencia fundamental de cada una de las sílabas. Igualmente, se calculó el número de sílabas por segundo de cada clase de emoción, con el fin de determinar la velocidad de habla. Además, se calculó el valor de la duración promedio de los fines de constituyentes textuales² y la duración de las sílabas que marcan inicio de frontera sintáctica.

Los patrones temporales de cada una de las sílabas fueron organizados en una hoja de cálculo con el fin de contrastar la duración correspondiente a cada sílaba, su posición en el sintagma y en el enunciado. Con los datos obtenidos se procedió a evaluar los resultados y contrastarlos con los reportados por los investigadores que han ahondado en el tema.

2.3. La velocidad del habla

Para el estudio específico de la velocidad de habla, entendida como *el número de sílabas divididas entre el tiempo total de locución (aquel que incluye el tiempo de*

² Para Blondet (2006), los constituyentes textuales pueden ser mayores (párrafos) o menores (cláusulas y constituyentes de cláusulas).

pausas), partimos de los hallazgos registrados por Blondet (2006), quien, para el español venezolano, propone una serie de rangos que permiten distinguir entre tres temporemas³ básicos: el temporema lento, el temporema normal y el temporema rápido. El primero se caracteriza por contar con un número de sílabas estimado entre las 3,49 y 4,1 sil/seg y por presentar distribuciones regidas por reglas como: mayor duración de sílabas en fines de constituyentes textuales, fines de constituyente textual mayor marcados doblemente (mayor duración silábica más pausa), y posibilidad de marcar fines de constituyente textual menor a través de pausa y de mayor duración silábica. El temporema normal está caracterizado por contar con un número de sílabas estimado entre las 4,7 y 5,3 sil/seg y por reflejar distribuciones regidas por reglas como: mayor duración de sílabas en fines de constituyente textual, fines de constituyente textual menores marcados ocasionalmente por pausa o a través de mayor duración silábica. Por su parte, en el temporema rápido presenta un rango estimado entre 6 y 7,1 sil/seg y muestra distribuciones regidas por reglas como: mayor duración de sílabas en fines de constituyentes textuales, fines de constituyentes textuales mayores marcados con pausa, y fines de constituyente textual menores marcados sólo con mayor duración.

2.4. El test de percepción

Un test de percepción se diseñó especialmente para la investigación. Este test se aplicó a 225 personas (134 mujeres y 91 hombres⁴), con una edad promedio de 26,4 años, todos hablantes de español venezolano. Durante la aplicación del test, las personas escucharon las 15 grabaciones del enunciado antes mencionado expresado con cada una de las emociones señaladas. En el test los participantes debían seleccionar de entre tres opciones posibles la emoción percibida en cada una de las grabaciones. Una vez aplicado el test se procedió a calcular el porcentaje de acierto para cada uno de los 15 ítems del test con el fin de determinar cuáles eran las emociones mejor y peor percibidas.

3. RESULTADOS

3.1. Análisis acústico: la melodía

Después de haber realizado el análisis acústico hemos encontrado que cada una de las emociones expresadas en las 20 grabaciones presentan características distintas. Los datos de los promedios obtenidos para cada uno de los parámetros acústicos se presentan en la Tabla 2.

³ Según Blondet (2006:2), el temporema es una unidad fonológica que regula el tempo de habla. “Es la instancia prosódica relacionada con el juego de duraciones y distribuciones de las sílabas y de las pausas para generar una velocidad específica”. El temporema es para Blondet (2006) una unidad equivalente al fonema.

⁴ No se tomaron en cuenta en esta investigación las diferencias de percepción según el sexo del oyente.

Tabla 2. *Parámetros acústicos de las declarativas neutras y de cada una de las emociones analizadas.*

	F0 (HZ)	INTENSIDAD (DB)	DURACIÓN TOTAL (SEG)
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

Según estos datos podemos afirmar que:

- 1) El dominio de la frecuencia fundamental parece tener una importancia particular para la codificación de la rabia, igualmente la intensidad tiene un rol vital. La rabia parece estar caracterizada por un incremento de la f0 y de la intensidad. Como veremos en la Tabla 3, en la rabia también se puede apreciar un incremento de la variabilidad y del rango de la f0.
- 2) La alegría, al igual que la rabia, presenta un incremento de la frecuencia fundamental y de la intensidad.
- 3) En la tristeza encontramos un descenso de la intensidad y de la f0.
- 4) 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.

Tabla 3. *Variabilidad melódica y temporal de las emociones y las declarativas neutras*

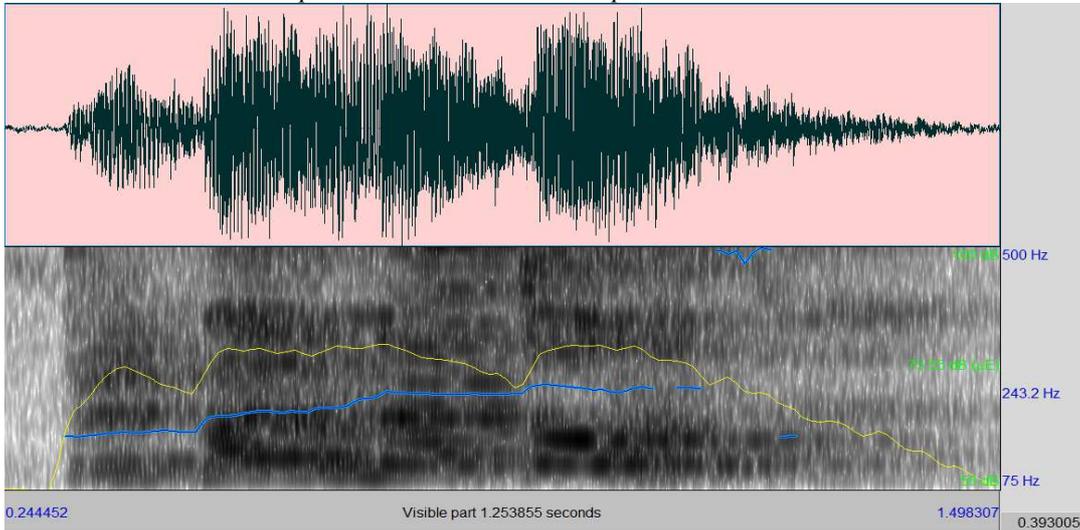
	FRECUENCIA FUNDAMENTAL			DURACIÓN		
	VARIABILIDAD (%)	TONO PROMEDIO (HZ)	DESVIACIÓN TÍPICA (HZ)	VARIABILIDAD (%)	DURACIÓN PROMEDIO (SEG)	DESVIACIÓN TÍPICA (SEG)
RABIA	10,19	240,32	24,49	10,19	1,42	0,0611
ALEGRÍA	10,3	202,83	20,91	7,41	1,08	0,0801
TRISTEZA	5,27	183,77	9,7	5,96	1,17	0,0698
NEUTRO	32,03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

Observamos que la rabia y la alegría presentan valores similares para la frecuencia fundamental y la intensidad. Distinto ocurre con la tristeza que presenta valores por debajo de los que presentan las otras dos emociones. Parecieran ser, entonces, la f0 y la intensidad los parámetros que permiten distinguir la tristeza de la rabia y la alegría. La distinción entre estas dos últimas emociones, sin embargo, viene dada por la duración. Se puede observar, por ejemplo, que la mayor duración silábica corresponde a la rabia (0,21656 seg), seguida de la tristeza (0,17579 seg) y la alegría (0,16548 seg)⁵. En otras palabras, la variabilidad o fluctuación melódica en el pretonema, diferencia la tristeza del resto de las emociones analizadas, mientras que la variabilidad temporal nos permite distinguir entre la rabia y la alegría, así como también distinguir cualquier tipo de emoción de las declarativas neutras (Cf. Tabla 3).

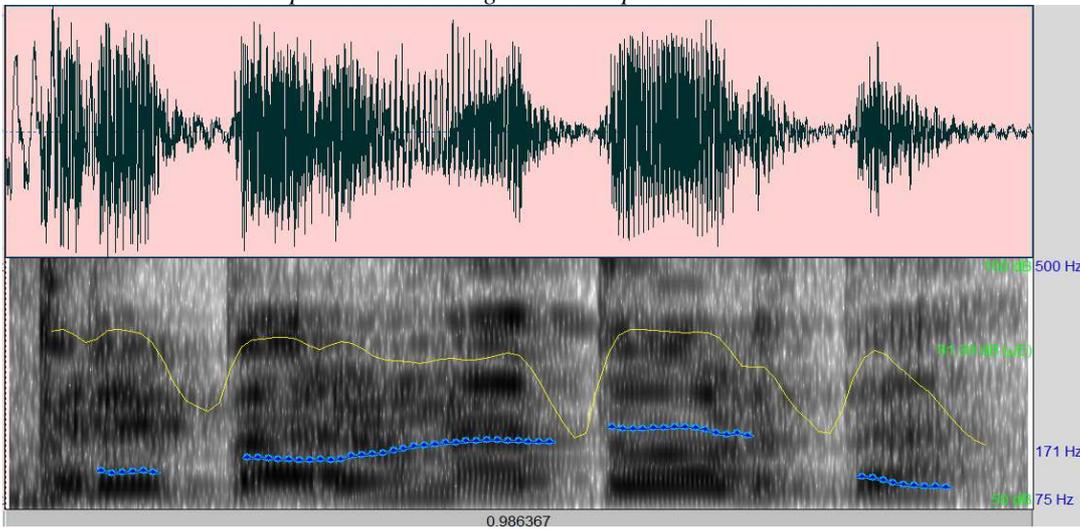
⁵ Estos datos parecen corroborar lo ya apuntado por Jiménez (1987), pues la duración, al igual que en la investigación de este autor, discrepa de los datos obtenidos para el inglés.

A continuación se presentan tres espectrogramas obtenidos con Praat correspondientes a 3 de las 15 grabaciones para ejemplificar los tres tipos de emociones analizadas:

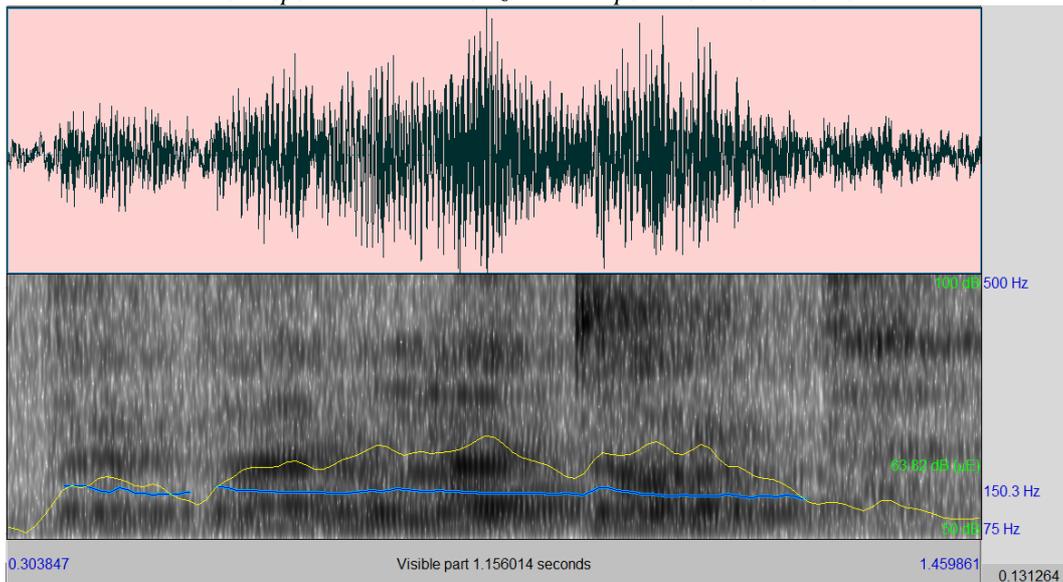
Espectrograma 1. *Curva de F0, curva de intensidad y espectrograma de uno de las grabaciones correspondiente a la rabia emitida por uno de los actores.*



Espectrograma 2. *Curva de F0, curva de intensidad y espectrograma de uno de las grabaciones correspondiente a la alegría emitida por uno de los actores.*



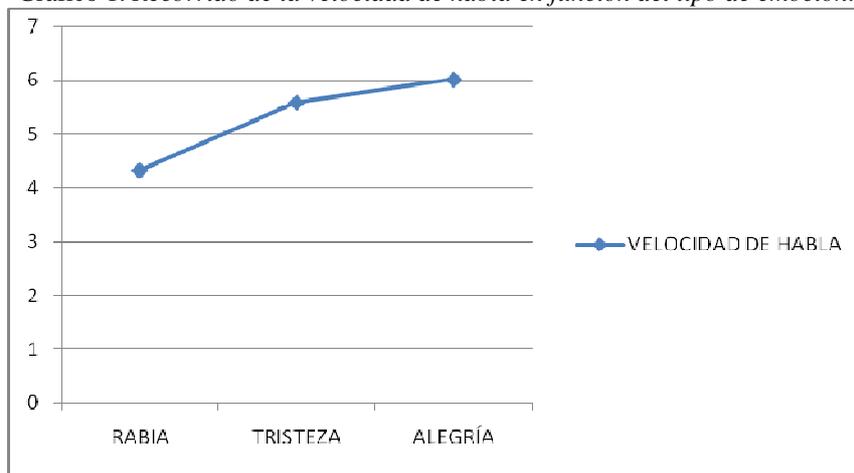
Espectrograma 3. Curva de F0, curva de intensidad y espectrograma de uno de las grabaciones correspondiente a la tristeza emitida por uno de los actores.



3.2. Análisis de la velocidad del habla y la duración silábica

En cuanto al número de sílabas por segundo utilizadas en la codificación de las emociones se pudo determinar que, como lo demuestra el Gráfico 1, esta asciende en función del tipo de emoción. En la rabia, por ejemplo, el número de sílabas por segundo es de 4,33, en la tristeza el número de sílabas por segundo asciende a 5,59 y en la alegría el número de sílabas por segundo es 6,01.

Gráfico 1. Recorrido de la velocidad de habla en función del tipo de emoción.

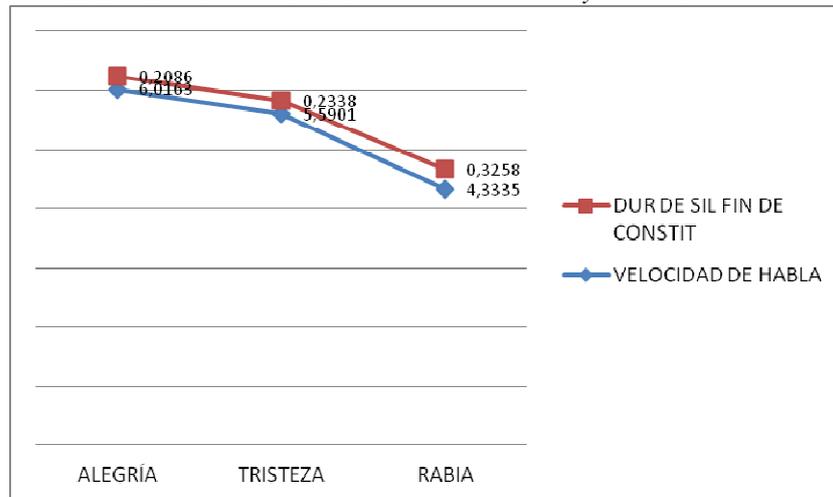


Si se comparan estos valores con los registrados por Blondet (2006) para el español venezolano, tenemos que la rabia se acercan en mayor medida a los rangos aportados para la velocidad de habla lenta, mientras que la tristeza, al igual que las declarativas neutras, se acercan más a los rangos aportados por la autora para la velocidad de habla normal, la alegría, por su parte, coincide con los valores aportados para la velocidad de habla rápida.

Por otra parte, si se comparan nuestros datos con los registrados para el francés por Beller et al. (2006) se observa una notable diferencia: velocidad lenta para la rabia y rápida para la alegría, en español venezolano, y velocidad más rápida para la rabia y menos rápida para la alegría en francés.

De igual forma, los datos evidencian una correlación entre el ascenso de la velocidad de habla y la duración de las sílabas que marcan el fin de los constituyentes textuales o de las fronteras sintácticas. Pues, mientras más cercana a lo lento parece ser la emoción, mayor parece ser la duración de la sílaba que marca el final del enunciado (0,3258 para la rabia). Y mientras más cercana a normal o lo rápido es la emoción menor parece ser la duración silábica (0,2338, 0,2086; para la tristeza y la alegría, respectivamente) (Cf. Gráfico 2).

Gráfico 2. Correlación entre velocidad de habla y duración silábica.



Esta correlación entre los elementos que marcan frontera sintáctica y la velocidad de habla corroboran que, como afirmaba Blondet (2006), cada temporema utiliza estrategias distintas para codificarse y entre estas estrategias la duración parece ser fundamental, pues permite al hablante desplazarse entre los distintos tipos de velocidades incluso cuando la extensión de los textos es relativamente breve.

En síntesis, proponemos una clasificación prosódica de las emociones basada en las siguientes características:

Tabla 4. Características prosódicas de las emociones y las declarativas neutras en el español meridiano

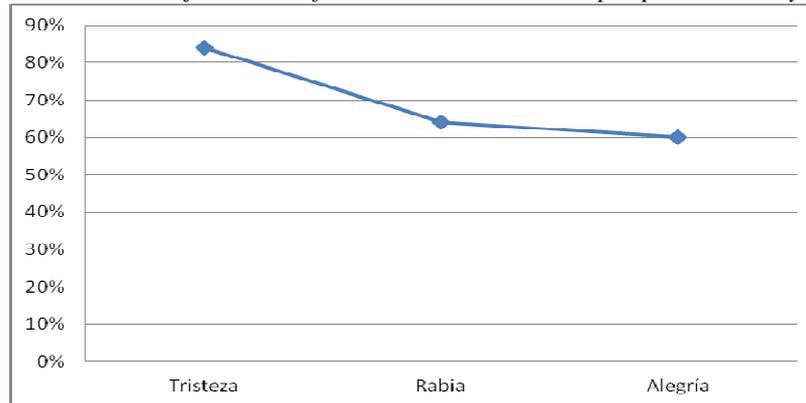
	F0 (HZ)	INTENSIDAD (DB)	DURACIÓN TOTAL (SEG)	DURACIÓN SILÁBICA (SEG)	VELOCIDAD DE HABLA (SIL/SEG)
RABIA	240,32	77,51	1,42	0,21656	LENTA
ALEGRÍA	202,83	74,81	1,08	0,16548	RÁPIDA
TRISTEZA	183,77	61,68	1,17	0,17579	NORMAL
NEUTRO	176,84	71,77	1,12	0,17149	NORMAL

Los datos acústicos y de velocidad de habla obtenidos para cada una de las emociones fueron analizados estadísticamente con el fin de determinar si las diferencias entre ellas son significativas o no. Para ello se utilizó la prueba Alfa de Cronbach. Esta prueba arrojó como resultado que la diferencia entre las emociones es significativa ($=,894$).

3.3. Análisis perceptivo

Desde el punto de vista perceptivo, hemos encontrado que la emoción mejor percibida por los oyentes es la tristeza (84%), seguida por la rabia (64%) y la alegría (60%)⁶ (Cf. Gráfico 3). Se obtuvo un porcentaje de precisión promedio de alrededor del 70%, que es aproximadamente cinco veces mayor de lo que se esperaría si los oyentes hubiesen respondido de forma aleatoria a los distintos ítems del test.

Gráfico 3. Porcentajes de identificación de las emociones por parte de los oyentes



Los datos arrojados por el test nos han permitido, igualmente, observar 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%)⁷. Estas no discriminaciones podrían obedecer, en ambos casos, a la similitud entre los valores de f0 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).

4. CONCLUSIONES

*La palabra puede disimular la emoción;
el tono la descubre.*
(Navarro Tomás, 1976:177).

Concluimos este análisis corroborando que los parámetros acústicos y la velocidad de habla funcionan para codificar y diferenciar emociones como la rabia y la tristeza de otras como la alegría.

⁶ Afirman Jiménez y Mallo (1988) en su estudio, al igual que nosotros, que la alegría es la emoción peor percibida. Scherer (2003), indica que la rabia es la emoción que mayor grado de exactitud obtiene en una gran cantidad de estudios revisados por él, seguida por la tristeza.

⁷ A pesar de que la media de reconocimiento está sobre el 70%, podríamos afirmar, junto con Scherer (2003), que los oyentes parecen tener problemas para identificar de forma inequívoca las emociones solamente a través en la voz. Según este investigador, la tristeza y la ira son generalmente los mejores reconocidos, seguido por el miedo. En varias investigaciones se ha determinado que contar con el soporte visual de la emoción ayuda en mucho a la percepción inequívoca de las emociones.

El análisis desde la producción de las distintas emociones y su percepción nos permite afirmar, a su vez, que más que existir una variable acústica que caracterice cada emoción, lo que realmente existe es un conjunto de variables que se coordinan de una determinada manera para caracterizar cada tipo de emoción. Dentro de las variables a considerar parece ser fundamental la duración, tanto promedio como interna de las sílabas, debido a que los distintos patrones de duración de las sílabas finales del enunciado permiten distinguir algunas de las emociones. Este mismo hecho lo afirma Fricke (2008) para el portugués.

Por su parte, la imposibilidad de distinguir claramente por medio de parámetros temporales y acústicos emociones como la alegría y la tristeza no hace más que corroborar la necesidad de partir, en el análisis de las emociones, del estudio de un conjunto de parámetros acústicos que vayan más allá del análisis temporal y acústico.

Uno de los objetivos de esta investigación era determinar si las emociones eran percibidas por los oyentes y en qué grado. A partir de los resultados obtenidos podemos responder afirmativamente a la primera parte de esta cuestión, ya que todas las emociones analizadas fueron reconocidas por encima del azar. Así mismo, los resultados nos indican que algunas emociones son significativamente mejor reconocidas que otras: es el caso de la tristeza. Explicar las razones de tales diferencias resulta, sin embargo, complicado y un tema poco abordado por otros investigadores. Nos atrevemos a afirmar que la percepción de las emociones puede variar de un oyente a otro, dependiendo de varios factores paralingüísticos (como el mismo estado anímico en el que se encuentre la persona) y socioculturales.

Este trabajo nos ha ayudado a entender la complejidad que representa el análisis de las emociones en el habla. Como resultado, creemos que no existe una única manera de etiquetar o clasificar las emociones a partir de la prosodia, varios deben ser los factores que se deben considerar para esta tarea.

5. AGRADECIMIENTOS

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.

6. REFERENCIAS BIBLIOGRÁFICAS

Banse, R., y Scherer, K. 1996. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70. 614-636.

Beller, G.; Hueber, T.; Schwarz D. y Rodet X. 2006. *Speech rates in French expressive speech*. Paris: Ircam, Institut de Recherche et de Coordination Acoustique/Musique.

Blondet, M. 2006. *Variaciones de la velocidad de habla en español: patrones fonéticos y estrategias fonológicas. Un estudio desde la producción*. Tesis doctoral. Mérida: Universidad de Los Andes.

Bouzon, C. 2004. *Rythme et structuration prosodique en anglais britannique contemporain*. Tesis Doctoral. Université Aix-Marseille I.

Fricke, A. 2008. Relating emotional content to speech rate in Brazilian Portuguese. En Barbosa, P. A., Madureira, S., and Reis, C. (Eds.) *Proceedings of 4th the Speech Prosody 2008 Conference*. Campinas, Brazil: Editora RG/CNPq.

Jimenez, F. 1987. *Marcadores emocionales en la conducta vocal*. Madrid: Universidad Autónoma de Madrid.

Jiménez, F. y Mallo, M. 1988. El reconocimiento de emociones a través de la voz. *Estudios de psicología*, 33-34. 31-52.

Johnstone, T. y Scherer, K. 2000. Vocal communication of emotion. En M. Lewis y J. Haviland (Eds.). *Handbook of emotions*. New York: Guilford Press. 220-235.

Muñoz, A. y Jiménez F. 1990. La expresión de la emoción a través de la conducta vocal. *Revista de Psicología General y Aplicada*, 43 (3). 289-299.

Scherer, K. 1982. Emotion as a process: function, origin and regulation. *Social Science Information*, 21. 550-570.

Scherer, K. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40, 227-256.

Scherer, K.; Johnstone, T. y Klasmeyer, G. 2003. Vocal expression of emotion. En Davidson, R.; Scherer, K. y Goldsmith, H. (Eds.). *Handbook of the affective sciences*. New York and Oxford: Oxford University Press. pp. 433-456.

Scherer, K. 2005. What are emotions? And how can they be measured? *Social Science Information*, 44, (4). 695-729.

Trouvain, J. 2004. *Tempo Variation in Speech Production. Implications for Speech Synthesis*. Tesis Doctoral. Saarland University.

Viola, I y Madureira, S. 2007. Voice quality and speech expressiveness. En *Proceedings of the Paralinguistic Speech07- between models and data*. Saarbrücken: DFKI Publication. 65-70.