

MODELIZACIÓN ACÚSTICA DE LA EXPRESIÓN EMOCIONAL EN EL ESPAÑOL.

Ángel Rodríguez Bravo, Patricia Lázaro, Norminanda Montoya, Josep M^a Blanco, Dolors Bernadas, Josep Manel Oliver y Ludovico Longhi.

Universidad Autónoma de Barcelona
Dto. De Comunicación Audiovisual y Publicidad
Edificio I, 08193 Bellaterra (Barcelona)
Arbravo@intercom.es

En esta comunicación se exponen los resultados del proyecto "*Modelización Acústica de la Expresión Emocional en el Español*" que ha sido financiado por la DGICYT (PB94-0732). Se parte de la construcción inicial de un corpus con 336 discursos en los que 8 actores simularon 7 emociones básicas (*alegría, deseo, rabia, miedo, sorpresa, tristeza y asco*), que fueron juzgados mediante tests de recepción por 1.054 sujetos experimentales. De esta primera etapa se seleccionaron los 34 discursos mejor valorados en el test para ser analizados y modelizados acústicamente. En el desarrollo del análisis acústico se manejaron 17 parámetros numéricos asociados a una perspectiva suprasegmental (análisis comparativo del contorno de F0 y de la presión acústica inter-locutor e intra-locutor). Se consiguió modelizar 6 de las 7 emociones estudiadas.

1. INTRODUCCIÓN.

Darwin vinculó la expresión sonora de las emociones al instinto o, cuanto menos, a conductas heredadas y, por tanto, que pueden aparecer en el individuo de modo totalmente involuntario (Cfr. DARWIN, 1872/1984). Desde ese momento, la expresión emocional ha tendido a ser considerada como el carácter del habla humana más claramente universal y transcultural. De hecho, el carácter transcultural del habla emocionada ha sido contrastado en diversas investigaciones muy posteriores, podemos citar por ejemplo los trabajos de KRAMER, 1964; BEIER y ZAUTRA, 1972; ALBAS, McCLUSKEY y ALBAS, 1976; McCLUSKEY, ALBAS y NIEMI, 1975; Van BEZOOIJEN, 1984. En todos estos estudios se explora ese supuesto carácter universal del sonido oral de las emociones y, efectivamente, se encuentra y se constata. No obstante, en estos trabajos se detectaron también elementos de influencia cultural que han desdibujado parcialmente los planteamientos darwinianos iniciales.

Nuestra investigación responde a la perspectiva híbrida que se apuntaba ya en los estudios transculturales que hemos citado más arriba. Este punto de vista que articula

la perspectiva "universalista" con la "culturalista" fue formalizado luego por Scherer. Este autor considera que la expresión oral de las emociones se constituye como un sistema analógico-vocal conectado con los mecanismos biológicos y fisiológicos del individuo, al cual se ha sobrepuesto después, a lo largo de la evolución, el sistema simbólico de la lengua (SCHERER, 1982). Así, en el estudio que presentamos hemos trabajado desde la doble perspectiva hipotética siguiente:

1. la voz sufre cambios acústicos que están motivados directamente por las alteraciones fisiológicas que se producen en el cuerpo humano cuando el individuo experimenta una emoción;
2. los rasgos acústicos de la voz producidos por las emociones resultan parcialmente modificados durante una locución por la influencia específica de la lengua que está utilizando el hablante emocionado.

No obstante, no nos parece adecuado diferenciar entre procesos lingüísticos y no lingüísticos, discriminando entre paralinguaje y extralinguaje en la línea de las propuestas hechas en su momento por LAVER y TRUDGILL, 1979; sino que nos hemos enfrentado al habla emocionada como a un sistema unitario que articula paralela y simultáneamente la influencia

cultural de la lengua con los mecanismos fisiológicos de la emoción, como un todo unívoco. Hemos buscado, pues, formas sonoras supra-segmentales que, con toda seguridad, articulan a la vez los caracteres sonoros propios de la lengua, con los rasgos acústicos que determina cada estado emocional.

Consideramos que esta forma de aproximación al problema del habla emocionada podría ser más fértil y reveladora que la orientación, hoy todavía dominante, que intenta caracterizar categorías discretas que permitan identificar de forma aislada emociones básicas y, por tanto, teóricamente independientes de la lengua (Cfr. FRICK, 1985; SCHERER, 1985, 1986; Van BEZOOIJEN, 1984).

Estudios más recientes tienden considerar que el número de parámetros que pueden caracterizar acústicamente a una emoción puede ser amplio, y trabajan simultáneamente con varios parámetros acústicos (F0, contorno tonal, duración, presión sonora, espectro, ...), no obstante, en el estudio del habla emocionada se sigue centrando la observación en fragmentos de habla sumamente cortos: una vocal, una palabra, o a lo sumo un único grupo fónico (Cfr. RUIZ, LEGROS y GUELL, 1990; RUIZ y LEGROS, 1994; LEINONEN, HILTUNEN, LINNANKOSKI y LAAKSO, 1997). A nuestro modo de ver, la observación de segmentos muy cortos del habla no permiten localizar y discriminar cuáles son los rasgos acústicos específicos de la emoción y cuál es el papel de la influencia de una determinada lengua sobre ellos.

Las emociones sufren una evolución dinámica en el tiempo, y su duración es limitada y variable (Cfr. EKMAN, 1983). En consecuencia, es muy difícil saber si un fragmento corto de habla seleccionado para un corpus ha coincidido exactamente con el momento de la expresión de la emoción, si hemos seleccionado una representación sonora de toda la evolución emocional completa o si, en realidad, solo la hemos atrapado parcialmente. Es necesario considerar, además, que el resultado de la interacción entre el carácter acústico de la lengua y el de la emoción depende de ambos a la vez. Es decir, que el resultado acústico

del habla emocionada, no solo depende de la evolución de la emoción sino que depende, también, del tipo de articulación, entonación, vocalización, estructura temporal, etc. específicos de la secuencia lingüística con la que esa emoción concreta coincide en cada momento. Así, la forma de garantizar una modelización eficaz de la expresión emocional del habla ha de ser, lógicamente, estudiando discursos orales completos. Una comparación posterior entre los resultados que puedan haberse obtenido utilizando este procedimiento con idénticos estados emocionales pero en lenguas distintas, posiblemente nos permitirá detectar con claridad cual es el papel de cada uno de los dos sistemas de estímulos (lingüístico y emocional) en las distintas formas acústico-emocionales del habla.

2. METODOLOGÍA.

2.1. Tests de recepción.

El trabajo que presentamos combina una serie de pruebas previas de recepción con el análisis acústico del habla emocionada. Esta es una metodología de trabajo que se inicia ya a finales de los 60 (Cfr. HUTTAR, 1967, 1968; WILLIAMS y STEVENS, 1972; Van BEZOOIJEN, 1984) y se basa esencialmente en comparar los resultados obtenidos en la discriminación perceptiva de sujetos que escuchan las voces emocionadas, con los resultados obtenidos discriminando las voces emocionadas a partir de procedimientos de análisis acústico. Puesto que en nuestra investigación, el objetivo no era encontrar parámetros discriminadores de las emociones sino su modelización sonora, hemos planteado la relación entre estos dos instrumentos ya clásicos en el estudio del habla emocionada de un modo bastante distinto al habitual. Se partió de un corpus extenso de discursos emocionales completos (336 discursos), que fueron escuchados y juzgados por 1.054 sujetos experimentales. Tras el análisis estadístico de los resultados, de los 336 discursos del corpus se seleccionaron solamente 34 para ser analizados y modelizados. En suma, no se utilizaron las pruebas de recepción para valorar la eficacia de los discriminadores

elegidos para el análisis acústico, sino que se recurrió al test perceptivo para seleccionar discursos que tuviesen todas las garantías de contener las formas acústicas emocionales que estábamos buscando.

2.2. Construcción del corpus.

El corpus inicial se preparó trabajando con 8 actores como informantes (4 hombres y 4 mujeres), a cada actor se le entregaron las correspondientes copias de los 2 textos portadores y se le solicitó que interpretase cada uno de los dos textos expresando las siguientes 7 emociones básicas: *alegría, deseo, asco, miedo, rabia, sorpresa y tristeza* con 3 grados de intensidad distintos (bajo, medio y alto). Es decir: (8 actores) x (2 textos) x (7 emociones) x (3 grados de intensidad) = 336 discursos en total. Con objeto de orientar a los actores se les proponía una frase concreta para sugerir cada grado de intensidad emotiva. Así, por ejemplo en el caso de la *alegría* las frases de sugerencia fueron: "te sientes contento", "estás entusiasmado" y "estás completamente eufórico". Antes de iniciar la sesión se solicitaba siempre a cada actor una interpretación neutra, es decir, sin emoción alguna.

La selección de las emociones básicas que se propusieron a los actores se apoyan en el acuerdo generalizado que hay entre los distintos estudiosos de la emoción respecto a que existe solo un número reducido de emociones básicas (Cfr. IZARD: 1977; DANTZER 1988: 29; EKMAN, 1983; REEVE, 1994:373). En lo que ya no existe acuerdo concreto es en el número exacto de éstas (que oscila entre 5 y 10) ni en los términos utilizados para denominarlas. Probablemente esta falta de acuerdo terminológico para denominar las emociones básicas siga siendo uno de los problemas centrales que dificulta la obtención de resultados homogéneos entre las distintas investigaciones sobre habla emocionada (Cfr. SCHERER, 1986).

Las grabaciones de las voces durante las interpretaciones se realizaron en los estudios de audio de la Facultad de Ciencias de la Comunicación de la Universidad Autónoma de Barcelona. La cadena de grabación utilizada fue la siguiente: micrófono "Senn-

heiser MD-421-1-4", mesa de mezclas "Tecnología Electrónica" y magnetófono "Revox- PR-99". Toda la cadena se calibró minuciosamente antes de cada grabación con una señal senoide pura de 1.000 a 100 dB, y las condiciones de calibrado fueron exactamente las mismas para todas las grabaciones, con objeto de que luego fuese posible desarrollar un análisis comparativo inter-locutor de la presión sonora.

2.3. Desarrollo de los tests de recepción.

Los tests de recepción se realizaron sometiendo cada una de las 336 interpretaciones emocionadas al juicio de dos grupos de más de 30 sujetos experimentales. Organizados en distintos grupos, se hizo escuchar a los sujetos experimentales una serie de 33 interpretaciones (la duración de cada interpretación oscilaba entre 20 y 40 segundos). Tras cada interpretación, se solicitaba mediante un test a cada oyente que respondiese a las tres cuestiones siguientes: 1) debía concretar qué emoción o emociones reconocía en cada voz; 2) tenía que asignar un grado de verosimilitud al locutor; y, por último, 3) debía especificar si había llegado a emocionarse, o no, escuchando la interpretación y, si la respuesta era afirmativa, concretar en qué grado. Este tipo de test nos permitió decidir con objetividad, es decir, con total independencia de los investigadores, qué interpretaciones contenían realmente informaciones acústicas concretas asociadas a las emociones. Tal como se ha comentado ya más arriba, en total intervinieron en el test 1054 sujetos experimentales, en su mayoría estudiantes.

Una vez realizado el análisis estadístico de los tests de recepción se adoptó el criterio de seleccionar las 4 o 5 interpretaciones de cada una de las emociones que habían obtenido un porcentaje de identificación y un grado de verosimilitud más altos. Así, las voces que luego fueron utilizadas para la modelización habían sido identificadas como portadoras de una expresión pura por más del 70 % de los sujetos que las habían juzgado. Respecto a la verosimilitud, el criterio adoptado fue que la media de verosimilitud obtenida fuera superior a 3 (las posibilidades de calificación en el test iban

de 1 a 7). Para valorar el alcance de estos resultados debe tenerse en cuenta que el test daba libertad a los jueces para clasificar las voces como portadoras de varias emociones combinadas, y que las voces seleccionadas para la modelización acústica fueron percibidas como cargadas de una emoción básica no combinada con ninguna por entre un 70 % y un 96 % de los sujetos que las valoraron.

2.4. Análisis acústico.

Tal como se ha mencionado más arriba, nuestro criterio de observación fué híbrido, y contemplaba tanto la búsqueda de categorías discretas vinculadas a aspectos muy concretos del estado emocional (como pueda ser, por ejemplo, la frecuencia fundamental asociada a la tensión laringea), como la observación de caracteres suprasegmentales como el contorno tonal o el de la intensidad.

Tras una serie de ensayos preliminares se preparó un protocolo que estudiaba sistemáticamente los siguientes parámetros en las 34 voces.-

Frecuencia fundamental:

- . F0 media,
- . rango de F0,
- . variabilidad de F0: [(F0 máx) - (F0 min)]

Presión sonora:

- . mediana durante el tiempo de fonación
- . rango en dB
- . variabilidad: [(dB max) - (dB min)]

Duraciones (ritmo):

- . tiempo total del discurso
- . tiempo de fonación
- . tiempo de pausas
- . nº de grupos fónicos
- . nº de pausas
- . duración de cada grupo fónico
- . duración de cada pausa
- . duración media del grupo fónico
- . duración media de la pausa
- . relación pausa-fonación en %
- . nº de sílabas por segundo.

A estos parámetros se añadieron en cada análisis dos tipos de representación gráfica:

A) Una representación global con todo el tiempo del discurso que mostraba un oscilograma, un análisis del pitch sobre un fondo escala de 0 a 400 Hz, y un análisis de la presión sonora sobre un fondo escala de 30 a 90 dB; y B) Una representación gráfica

parcial conteniendo sólo de 1 a 3 grupos fónicos, pero que mostraba exactamente los mismos elementos de análisis que la representación global.

El instrumento utilizado para estos análisis fue un CSL-4300B y las voces se muestrearon a 16.000 m./seg. No obstante, cuando los resultados del análisis (especialmente los del pitch) resultaban confusos se utilizó también el prototipo de analizador "ANETO" cedido a nuestro laboratorio por el "Grupo de Procesamiento de Señal y Comunicaciones" del Dto. de Teoría de la Señal y Comunicaciones de la U.P.C.

Finalmente, se realizó un análisis espectral de todos los discursos a partir de la señal analógica generada por el propio CSL-4300B (definición de 0 a 8000 Hz), pero esta vez utilizando un equipo BRÜEL 2143. Este último analizador nos permitía trabajar en bandas de octava, posibilidad de la que no dispone el CSL. Para esta última etapa el equipo se configuró en 1/3 de octava, analizando toda la banda pasante. La toma de muestras se realizó capturando 32 espectros promediados acumulativamente, generándose un espectro cada segundo. Es decir cada espectro acumulaba estadísticamente (haciendo una media aritmética) toda la información anterior, con lo que el último espectro, obtenido al final de la emisión sonora analizada, expresaba un espectro medio del conjunto de la locución.

3. RESULTADOS OBTENIDOS.

3.1. El concepto de estado promedio.

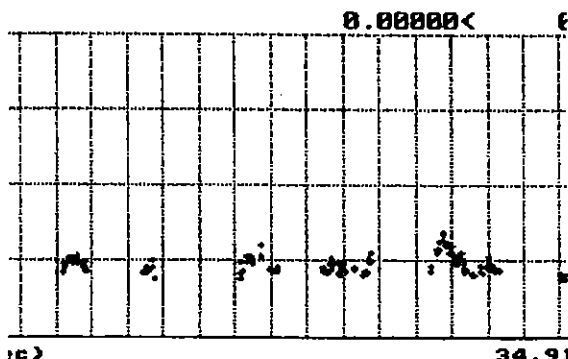
Como primer resultado podemos afirmar que el referente adecuado para observar los parámetros acústicos asociados a las emociones de un locutor es lo que hemos denominado como el estado-promedio. Definimos este estado como la media aritmética que se obtiene para un mismo locutor con los datos de cada parámetro concreto. Hemos resuelto, pues, el problema de la variabilidad expresiva en el análisis acústico del habla comparando los datos obtenidos en los distintos estados emocionales, con su propio estado-promedio. De ese modo, al poder observar las desviaciones

acústicas de cada voz respecto a su *estado-promedio*, queda establecido un criterio común de referencia intra-locutor que, luego, permite comparar entre sí las voces de locutores distintos.

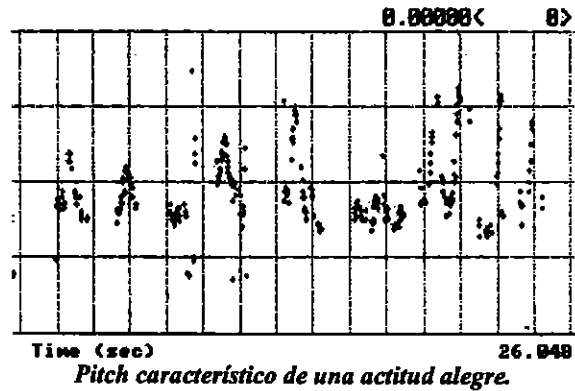
3.2. Características globales del habla emocionada.

Hemos comprobado que en la dinámica global de las estructuras prosódicas del habla emocionada, la estructura acústica que caracteriza una emoción solo se manifiesta completa y con total claridad en algunos de los grupos fónicos del discurso. Y que esto ya resulta suficiente para que un receptor identifique el estado emocional del locutor. Asimismo, hemos comprobado que mientras el contorno entonativo y de intensidad característicos de la emoción pueden mostrarse solo en algunos grupos fónicos, la estructura rítmica asociada a un estado emocional tiende a manifestarse a lo largo de la totalidad del discurso.

Hemos observado, también, que la dinámica entonativa característica del español evoluciona en el tiempo en forma de cresta o diente de sierra, ajustándose los máximos tonales a los puntos del grupo fónico correspondientes a las vocales tónicas y los mínimos a las átonas, y que esta cresta se ve alterada por los estados emocionales aumentando o disminuyendo su variabilidad. En los 2 gráficos que se muestran a continuación puede observarse con claridad como el mismo texto, al ser interpretado con una actitud triste, o con una actitud alegre, se ve radicalmente alterado en su contorno tonal, aumentando la variabilidad de su cresta en diente de sierra prácticamente al triple.



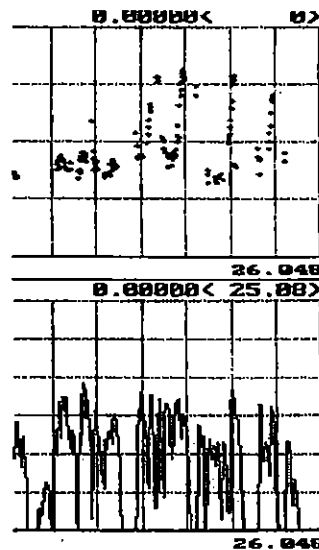
Pitch característico de una actitud triste.



3.3 Sobre el modelo acústico asociado a cada emoción.

En esta investigación hemos conseguido modelizar acústicamente la expresión emocional para 6 de las 7 emociones básicas estudiadas. A continuación expondremos cuáles son los rasgos fundamentales de cada uno de estos modelos acústico-emocionales:

ALEGRÍA.- Entonación: esta emoción produce un aumento global del tono medio que oscila entre el 10 y el 50 % según el grado de intensidad emocional; la estructura en diente de sierra se configura con grandes inflexiones tonales ascendentes-descendentes a partir del tono medio que pueden llegar a una desviación del 100 al 120 %, en especial al inicio del grupo fónico y luego van disminuyendo, eso supone un aumento de la variabilidad tonal respecto al *estado-promedio* del locutor en torno a un 120 %; estas inflexiones tonales se producen entre 2 y 3 veces por segundo.



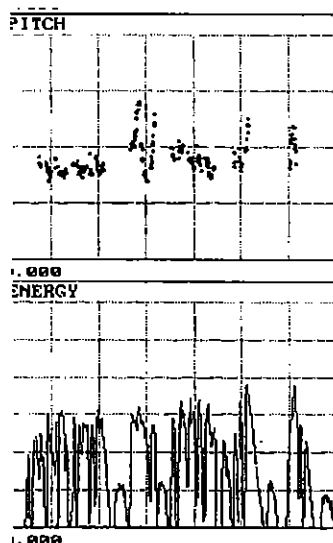
Análisis del pitch (arriba) y de la intensidad (abajo) de un grupo fónico característico de la ALEGRÍA.

Intensidad: tiende muy claramente a mantenerse estable durante todo el discurso.

Ritmo: Se produce una disminución del 20 % en la duración de las pausas.

DESEO.- Entonación: el tono medio baja en torno a un 10 %; la variabilidad tonal se reduce entre un 5 y un 10 %; las variaciones de la estructura en diente de sierra aparecen mucho más lentamente que en la alegría, llegando a durar cada inflexión ascendente-descendente unos 2 segundos.

Intensidad: a lo largo del grupo fónico se produce una gran caída regular de intensidad que puede llegar hasta los 25 dB; el final de cada grupo fónico se convierte casi sistemáticamente en una fuerte y expresiva espiración.



Análisis del pitch (arriba) y de la intensidad (abajo) de un grupo fónico característico del DESEO.

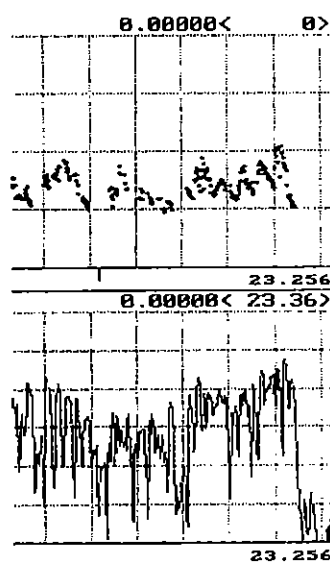
Ritmo: La fragmentación del discurso aumenta en un 20 %, produciéndose a la vez una reducción del 10 % y del 15 % respectivamente en el tiempo de los grupos fónicos y de las pausas; así, el deseo hace crecer la duración global del discurso aumentando la cantidad de pausas, no obstante, acorta cada uno de los fragmentos sonoros que lo componen.

RABIA.- Entonación: la estructura tonal en diente de sierra tiene unas variaciones que pueden oscilar entre 20 y 80 Hz. según la intensidad de la emoción, con una orientación global horizontal. **Intensidad:** el

grupo fónico característico tiene una intensidad claramente ascendente desde el inicio al final, subiendo entre 5 y 10 dB.

Ritmo: la rabia tiende muy claramente a disminuir la fragmentación del discurso oral, reduciéndose el número de pausas en un 25 %, a la vez, acorta la duración absoluta de cada pausa entre un 7 y un 8 %, lo que reduce la duración global del discurso.

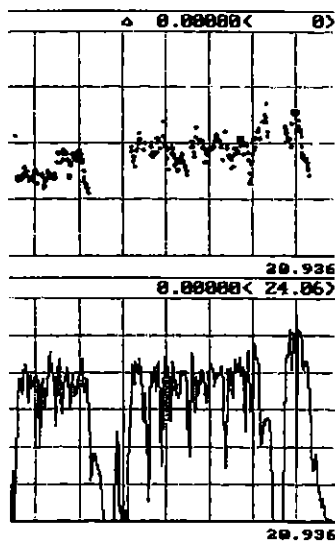
Timbre: el rasgo acústico más característico de la rabia parece ser un aumento global medio entre 10 y 15 dB. de la energía sonora en las bandas de 500-630 Hz. y de 2000-2500 Hz., a lo largo de todo el discurso.



Análisis del pitch (arriba) y de la intensidad (abajo) de un grupo fónico característico de la RABIA.

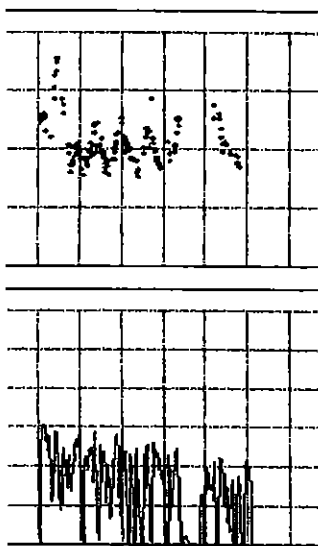
MIEDO.- Entonación: el miedo hace aumentar el tono medio de la voz de un 5 a un 10 %, y reduce la variabilidad tonal también muy ligeramente (un 5 %); la estructura entonativa que lo caracteriza es un diente de sierra globalmente ascendente, oscilando este ascenso global entre 50 y 100 Hz. en los grupos fónicos emocionalmente más relevantes. **Intensidad:** es globalmente ascendente a lo largo de todo el grupo fónico, siendo el ascenso de unos 10 dB.

Ritmo: el miedo acelera el ritmo, reduciendo el tiempo de duración global del discurso en un 35 %, de esa reducción 2/3 corresponden a la fonación y solo 1/3 a las pausas; la disminución del tiempo de fonación se consigue aumentando la velocidad de locución entre 1,5 y 2 sílabas por segundo, y la de las pausas reduciendo la fragmentación en un 25 %.



Análisis del pitch (arriba) y de la intensidad (abajo) de un grupo fónico característico del MIEDO.

SORPRESA.- Entonación: hace subir el tono medio de la voz de un 10 a un 15 % y aumenta la variabilidad tonal entre el 15 y el 35 %; la estructura entonativa es similar a la de la alegría, con un diente de sierra de grandes inflexiones ascendentes-descendentes que llegan a los 100 Hz., con el carácter diferencial de que finalizan en forma ascendente en grupos fónicos alternos (uno sí uno no, aproximadamente).

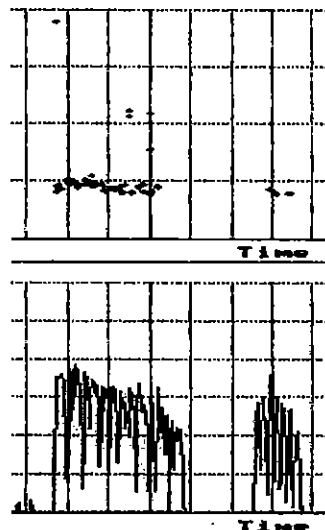


Análisis del pitch (arriba) y de la intensidad (abajo) de un grupo fónico característico de la SORPRESA.

Intensidad: la sorpresa aumenta la intensidad media entre 3 y 5 dB. y configura una fuerte caída de intensidad regular global en los grupos fónicos con final tonal descendente de entre 15 y 20 dB. **Ritmo:** el

tiempo de duración del discurso tiende a disminuir en torno a un 10 % al aumentar la velocidad de locución entre 0,5 y 1 sílaba por segundo.

TRISTEZA.- Entonación: reduce el tono medio entre un 10 y un 30 % y la variabilidad tonal entre el 30 y el 50 %, según el grado de intensidad emocional; la estructura entonativa característica de la tristeza es especialmente homogénea en el tono, éste tiende a ser muy bajo y la oscilación en diente de sierra es casi nula; la orientación global del tono es ligeramente descendente, pudiendo tener una caída global máxima de unos 50 Hz. a lo largo todo el grupo fónico.



Análisis del pitch (arriba) y de la intensidad (abajo) de un grupo fónico característico de la TRISTEZA..

Intensidad: la tristeza hace bajar la intensidad media de la voz entre un 10 y un 25 %; en los grupos fónicos más característicos produce un descenso progresivo de la intensidad muy acusado, que puede oscilar entre 20 y 30 dB. **Ritmo:** esta emoción ralentiza mucho el discurso aumentando su duración global entre un 20 y un 50 %; ese aumento se debe básicamente al alargamiento de las pausas, cada una de ellas puede crecer del 50 al 150 %; la velocidad de locución se hace también algo más lenta, reduciéndose entre 0,3 y 0,75 sílabas por segundo; esta ralentización no alarga los grupos fónicos sino que se traduce en el aumento de un 10 % en la fragmentación del discurso.

ASCO.- Esta es la única emoción que no ha podido ser modelizada acústicamente. La razón es que el "asco" nunca fue identificada auditivamente por más de un 50 % de los sujetos experimentales que juzgaron las voces que intentaban expresarlo. Lógicamente, esa gran diferencia con el resto de las emociones (todas consiguieron identificaciones por encima del 90 %) nos llevó a concluir que esta era una emoción muy poco reconocible en el sonido de la voz y, en consecuencia, que carecía de sentido intentar su modelización.

4. CONCLUSIÓN.

Con esta investigación se está proponiendo un nuevo modelo metodológico para el estudio de la variabilidad expresiva del habla que constaría de 4 partes bien diferenciadas:

1. *Construcción de un corpus amplio de discursos de habla emocionada.*
2. *Validación y selección del corpus mediante tests de recepción.*
3. *Análisis y modelización acústica del corpus.*
4. *Implementación de los resultados en un sistema de síntesis y contrastación definitiva mediante tests de recepción.*

El paso siguiente a esta investigación deberá ser, entonces, la contrastación de los modelos acústicos obtenidos, implementándolos en alguno de los sistemas de síntesis de habla en español. En esta última fase se construiría un nuevo corpus, ahora sintético, que será de nuevo sometido a tests de recepción. Estamos proponiendo, en suma, un procedimiento que, a pesar de que incluye y utiliza toda la instrumentación tecnológica actual, está muy próximo al funcionamiento real de los lenguajes naturales; y del cual se podrán obtener, finalmente, parámetros adecuados y eficientes para representar la información del discurso oral en cualquiera de los ámbitos de la ingeniería del lenguaje.

BIBLIOGRAFÍA:

- ALBAS, D.C.; McCLUSKEY, K.W. y ALBAS, C.A.: "Perception of the emotional content of speech: a comparison of two canadian groups". *Journal of Cross Cultural Psychology*, 7, 481-490, 1976.
- BEIER, E.C. y ZAUTRA, A.J.: "Identification of vocal communication of emotions across cultures". *Journal of Consulting and Clinical Psychology*, 39-166, 1972.
- BEZOOIJEN, R. Van: *Characteristics and recognizability of vocal expressions of emotion*. Dordrecht, Foris Publications, 1984.
- DARWIN, Ch.: *La expresión de las emociones en los animales y en el hombre*. Alianza Ed. Madrid, 1984.
- DANTZER, R.: *Las emociones*. Paidós, Barcelona, 1989.
- EKMAN: "An Argument for Basic Emotions", *Cognition and Emotion*, 6, pp. 1, 1983.
- FICK, R.W.: "Communicating emotion: the role of prosodic features". *Psychological Bulletin*, 97, 412-429, 1985.
- HUTTAR, G.L.: "Some relations between emotions and the prosodic parameters of speech". Monograph 1, Speech Communications Lab. Inc., Santa Bárbara, 1967.
- HUTTAR, G.L.: "Relations between prosodic variables and emotions in normal american english utterances". *Journal of Speech and Hearing Research*, 11, 481-487, 1968.
- IZARD, C. (ed.): *Measuring Emotions in Infants and Children*. Cambridge University Press, 1982.
- KRAMER, E.: "Judgments of personal characteristics and emotions from nonverbal properties of speech". *Psychological Bulletin*, 60, 408-420, 1963.
- LAVER, J. y TRUDGILL, P.: "Phonetics and linguistics markers speech". En K.R. SCHERER y H.GILES Eds., *Social Markers in Speech*. Cambridge, Cambridge University Press, 1979.
- LEINONEN, L. HILTUNEN, T. LINNANKOSKI, I. y LAAKSO, M.: "Expression of emotional-motivational connotations with a one-word utterance". *J. Acoustic Soc. Am.*, 102 (3), 1853-1863, 1997.
- McCLUSKEY, K.W.; ALBAS, D.C. y NIEMI, R.R.: "Cross cultural differences in the perception of the emotional content of speech; a study of development of sensitivity in canadian and mexican children". *Developmental psychology*, 11, 551-555, 1975.
- RUIZ, M.S.R. LEGROS, M.S.C. y GUELLA, A.: "Voice analysis to predict the psychological or physical state of a speaker". *Aviation Space, and Environmental Medicine*, 266-271, 1990.
- RUIZ, R. y LEGROS, C.: "The cumulative spectral probability diagram: theory and experiments". *Acta Acústica*, 4, 215-222, 1994.
- SCHERER, K.R.: "Methods of research on vocal communication: paradigms and parameters". en K.R. SCHERER y P.EKMAN Eds., *Handbook of Methods in non verbal behavior research*. Cambridge University Press, 1982.
- SCHERER, K.R.: "Vocal affect signaling: A comparative approach" en J.S. ROSENBLATT, C. BEER, M.C. BUSNEL y P.J.B. SLATER Eds., *Advances in the Study of Behavior* (Vol. 15). New York, Academic Press, 1985.
- REEVE, J.M.: 1994 *Motivación y emoción*. McGraw-Hill/Interamericana de España, S.A. Aravaca (Madrid).
- SCHERER, K.R.: "Vocal affect expression: a review and a model for future research". *Psychological Bulletin*, 99, 143-165, 1986.
- WILLIAMS, C. y STEVENS, K.N.: "Emotions and speech: some acoustical correlates". *J. Acoust. Soc. Am.* 52, 1238-1250, 1972.