

El habla emocionada en la comunicación audiovisual: una nueva metodología de análisis¹

ÁNGEL RODRÍGUEZ, PATRICIA LÁZARO, NORMINANDA
MONTTOYA, JOSEP MARÍA BLANCO,
DOLORS BERNADAS, J. MANEL OLIVER Y LUDOVICO LONGHI
Universidad Autónoma de Barcelona



Resumen

En este artículo se expone la necesidad de nuevas herramientas metodológicas para el análisis de la comunicación audiovisual y al hilo de la investigación expuesta, se presenta, también, un nuevo instrumento de estudio: el "método de análisis instrumental".

La investigación que exponemos muestra esta línea de trabajo experimental y estudia y formaliza el efecto de los distintos estados emocionales en la voz de un locutor. Hemos construido un corpus con 332 interpretaciones orales de 8 actores. Después se realizaron test de recepción con más de un millar de sujetos experimentales para objetivar estadísticamente las interpretaciones que expresaban adecuadamente cada emoción. Posteriormente, analizamos las voces "óptimas" mediante distintos procedimientos acústicos, y por último, se estudiaron las correlaciones entre las valoraciones de los sujetos y los modelos acústicos obtenidos, para descubrir cuáles son los arquetipos formales de la voz con los que se consiguen expresar cada una de las emociones.

Palabras clave: Voz, emoción, análisis acústico, percepción del habla.

Touched speech in audiovisual communication: A new analysis methodology

Abstract

This article explains the need of new methodological tools for the analysis of audiovisual communication and, in the context of the expressed research, a new instrument of study is also introduced: the «method of instrumental analysis».

The research we present shows this way of experimental work and studies and formalizes the effect of the different emotional shapes in a speaker's voice. We have build a corpus with 332 oral interpretations by 8 actors. After that we performed reception tests with over one thousand experimental subjects, to objetivate statistically the interpretations that expressed adequately each emotion. Later, we analyzed the "optimal" voices by means of different acoustic procedures, and lastly we studied the correlations between the subjects' judgment and the acoustic models obtained, to find out which are the formal archetypes of the voice with which we achieve to express several emotions.

Keywords: Voice, emotion, acoustic analysis, speech perception.

Que el habla es un recurso nuclear y fundamental en la comunicación audiovisual es indiscutible. Sin embargo, el número de estudios dedicados a la expresión hablada en el ámbito de la cultura audiovisual es escasísimo. Prácticamente la totalidad de los productos audiovisuales de cine, radio y televisión utilizan el sonido de la voz humana como un recurso expresivo básico; en cambio, es extremadamente difícil encontrar investigaciones en torno a la comunicación audiovisual que tomen la expresión oral como un tema central.

De hecho, en la literatura sobre comunicación audiovisual sólo es posible localizar unos pocos capítulos con referencias muy generales al sonido del habla. Y estos se encuentran siempre en el marco de trabajos mucho más generales. Es el caso de estudios como los de Hills (1981), Mcleish (1981), Muñoz y Gil (1986), Merayo (1992), Balsebre (1994), Cebrian (1994), etc. Ciertamente, esta lista no es exhaustiva, pero sí que refleja con precisión el tipo de aproximaciones a la expresión oral que han hecho tradicionalmente los estudiosos de la comunicación audiovisual. En cambio, se han derramado auténticos ríos de tinta estudiando y explicando cómo deben ser los textos destinados a la locución.

¿A qué se debe esa grave ausencia? ¿Por qué tendemos a ocuparnos sólo de la dimensión escrita del habla cuando en la fenomenología audiovisual el carácter sonoro es absolutamente primordial?

A nuestro modo de ver, esta importante carencia tiene una explicación metodológica. Resulta muy fácil fijar sobre un papel los contenidos léxicos y la estructura sintáctica del habla, utilizando simplemente su transcripción escrita; en cambio, analizar la expresividad sonora del habla se convierte siempre en una tarea bastante complicada. Es cierto que podemos grabar el sonido de la voz en un soporte magnético y repetirlo de forma técnicamente idéntica todas las veces que sea necesario; pero, aun así, los matices expresivos del texto oral seguirán sin poder ser fijados y en cada reproducción fluirán de nuevo en el tiempo escapando a toda capacidad de observación directa. En suma, un problema básico para el estudio del habla en los discursos audiovisuales es la necesidad de instrumentos metodológicos que nos permitan observar y objetivar cómo son y qué efecto comunicativo producen determinadas formas sonoras de la voz que evolucionan en el tiempo.

De hecho, este problema metodológico puede ser planteado de forma muy clara como una doble necesidad que, en realidad, también subyace en la mayor parte de las investigaciones vinculadas a los fenómenos comunicativos audiovisuales:

1. Necesitamos instrumentos de análisis que nos permitan objetivar las características de determinadas formas sonoras y visuales que evolucionan en el tiempo.
2. Necesitamos métodos de control que nos permitan objetivar el carácter de las percepciones que experimentan los sujetos expuestos a determinadas formas sonoras y visuales que evolucionan en el tiempo.

Toda comunicación audiovisual produce y organiza formas acústicas o lumínicas (también ambas a la vez) constituidas por decenas de variables, que evolucionan en el tiempo e interactúan entre sí como un sistema expresivo unitario del que sabemos muy poco. Y los instrumentos metodológicos habituales en las ciencias sociales no permiten determinar con precisión cuáles son estas variables y cuál es exactamente el papel expresivo de cada una de ellas. Es decir, no resuelven un problema tan esencial para la comunicación audiovisual como la objetivación de los mecanismos que actúan en el reconocimiento de formas. Vemos, pues, que las dificultades metodológicas que comporta en concreto el estudio de la expresión oral, son, también, extrapolables a muchos de los problemas de

conocimiento que siguen todavía abiertos en el marco de la comunicación audiovisual. Problemas para los que resultaría muy útil incorporar métodos² y conocimientos de otras disciplinas que ya han conseguido resolver cuestiones comunes o muy similares a las nuestras desde perspectivas sustancialmente distintas.

Nos estamos refiriendo a instrumentos desarrollados en los últimos 20 años por la física, la informática y la ingeniería de las comunicaciones, que permiten medir y comparar con precisión las distintas dimensiones acústicas y lumínicas, obtener de ellas datos numéricos y diseñar métodos de estilización para la obtención de modelos formales. Y nos referimos, también, a los instrumentos desarrollados en el marco de la física médica y la psicología de la percepción que facilitan el control objetivo de los fenómenos físicos que se desencadenan en el interior del ser humano.

Pero revisemos ahora el caso concreto de las emociones en el habla como un objeto de estudio que precisa de estos nuevos planteamientos metodológicos que exponíamos en el apartado anterior.

¿POR QUÉ LAS EMOCIONES?

Los objetivos de nuestra investigación parten de una vieja necesidad de conocimientos sobre la expresión oral. El saber concreto sobre cómo son y cómo deben construirse determinados matices expresivos del sonido del habla: agrada-bilidad, belleza, emociones, seguridad, tensión, ... etc. ha sido siempre muy escaso e intuitivo; estando, además, casi exclusivamente restringido al ámbito de la formación dramática. La comunicación audiovisual ha heredado directamente esa necesidad, y los modernos medios de difusión masiva, al revitalizar y multiplicar el uso de la expresión oral, han hecho patente la necesidad de un nuevo saber científico y técnico sobre lo oral. No sólo se precisa ese nuevo conocimiento para la formación de locutores, presentadores, periodistas audiovisuales, etc.; sino que el propio desarrollo tecnológico demanda cada vez más esos nuevos conocimientos. Todos los procesos de automatización sonora del habla precisan de nuevos conocimientos sobre la expresividad oral: ¿Cómo debe hablar un sintetizador de voz para que resulte natural? ¿De qué modo es posible diferenciar un estilo de locución de otro cuando se sintetiza un mismo texto? ¿Cómo puede identificar un sistema de reconocimiento automático de voz a un locutor de otro?

Así, el objetivo central de nuestro proyecto, desde un punto de vista general, era conseguir una ampliación de los conocimientos sobre las formas sonoras con las que se construye la expresividad oral. Es decir, queríamos llegar a saber, por ejemplo, por qué al leer dos veces un mismo texto, la primera locución puede llegar a producir sensación de agresividad y la segunda sensación de tristeza. Queríamos saber cuál era el modo de controlar esto: qué dimensiones acústicas de la voz había que cambiar para conseguirlo, y cuál era la forma de hacerlo. Como ocurre siempre que se profundiza lo suficiente en un objeto de estudio, éste se revela mucho más amplio de lo previsto y es necesario volver a acotarlo de nuevo para hacerlo más asequible. Este proceso lógico de necesidad de reducción de nuestro objeto de estudio es el que nos lleva por fin a las emociones.

Decidimos centrar nuestro estudio sobre la variabilidad expresiva del habla en las emociones por dos motivos muy básicos: 1) es un fenómeno comúnmente aceptado tanto en ámbitos populares y profesionales como científicos que las emociones tienen capacidad de influencia en los sonidos del habla; 2) la psicología define las emociones como fenómenos vinculados a alteraciones fisiológicas, y por tanto, como objetivables metodológicamente.

Para desarrollar una investigación que pretendía describir con rigor cómo es el sonido de cada una de las emociones en el habla, tuvimos que sumergirnos de lleno en esa doble necesidad que se planteó en el primer apartado. Por una

parte, necesitábamos instrumentos de análisis que posibilitaran la comparación objetiva del sonido de unos textos orales con otros. Y, por otra, precisábamos métodos de control que nos permitiesen diferenciar con objetividad qué estado emocional estaba realmente asociado a cada uno de los textos orales que pensábamos estudiar.

En nuestro caso concreto, la fonética experimental y la psico-acústica aportaron los instrumentos para la comparación objetiva entre las formas sonoras de las distintas voces emocionadas. Y la psicología de la percepción junto a la psicología de las emociones nos proporcionaron los conocimientos previos y el método de control para poder asociar cada voz a un estado emocional concreto. Se daban, pues, una serie de condiciones que nos permitían aplicar en nuestra investigación lo que vamos a denominar a partir de ahora como una *“metodología de análisis instrumental”*. Es decir, una metodología de análisis formal apoyada estrictamente en instrumentos de medida y que persigue de manera sistemática la contrastación experimental de todas sus hipótesis.

ETAPAS PARA UN ESTUDIO SOBRE LA EXPRESIVIDAD SONORA ORAL

La hipótesis general de la investigación fue la siguiente:

Las emociones vividas por un locutor influyen en las formas sonoras de su discurso oral alterándolas de forma homogénea, coherente y gradual. Las citadas alteraciones dependen directamente del tipo de emoción y del grado de intensidad de ésta.

Si la afirmación de la hipótesis era cierta, podríamos construir un corpus con distintas locuciones realizadas a partir de estados emocionales diferentes y, luego, analizar acústicamente estas locuciones para localizar en ellas el modelo sonoro que se configura en cada estado emocional.

Lógicamente, el planteamiento anterior expresa sólo la filosofía global de nuestro estudio. En primer lugar era necesario probar que realmente las emociones alteraban el sonido de la voz; y que esas alteraciones actúan luego como índices expresivos, es decir, que son reconocibles por cualquier receptor. Si podíamos comprobar esto, la consecuencia sería que, efectivamente, era posible realizar una formalización acústica del habla emocionada. Y, en segundo lugar, necesitábamos garantizar que las voces que íbamos a analizar expresaban realmente las emociones que deseábamos estudiar con la suficiente diversidad. En suma, debíamos someter a nuestro corpus a una serie de controles previos antes de proceder a su análisis instrumental. Así, para poder completar todo este proceso de control y análisis, el desarrollo de la investigación se realizó en las 5 etapas siguientes:

1. Diseño, producción y grabación del corpus.
2. Desarrollo de tests de recepción con todos los mensajes del corpus.
3. Análisis estadístico para la contrastación de las hipótesis generales y la selección de los mensajes comunicativamente “óptimos”.
4. Análisis instrumental (acústico) de los mensajes “óptimos”, obtención de datos y extracción de modelos formales.
5. Estudio de las correlaciones entre las etapas 3 y 4 y obtención de conclusiones.

Diseño, producción y grabación del corpus

Antes de iniciar una recopilación de voces emocionadas debíamos plantearnos de qué modo íbamos a manejar experimentalmente el fenómeno de las emociones y cómo orientaríamos y estableceríamos métodos de control con los locutores que nos ayudarían a construir el corpus.

Tras una revisión de la literatura sobre psicología de las emociones, llegamos a la conclusión de que existe un acuerdo ampliamente generalizado entre los psi-

cólogos en concebir la emoción como una reacción orgánica compuesta por tres componentes: 1) *experiencia subjetiva*, 2) *conducta expresiva*, y 3) *respuesta fisiológica* (Reev, 1994, p. 20). Así, las emociones no son exclusivamente experiencias subjetivas, sino que implican, también, cambios anatómicos, neurofisiológicos y endocrinos involuntarios, que se desencadenan al experimentar la emoción (Izard, 1982). Ese carácter fisiológico y no voluntario da a la emoción un rasgo diferencial muy claro, por ejemplo, respecto al concepto más vago y difuso de *sentimiento*, en tanto que con los sentimientos los individuos no experimentan modificaciones somáticas (Dantzer, 1989, p. 29); convirtiendo así las emociones en un fenómeno relativamente fácil de identificar en una investigación.

Estábamos, pues, frente a un fenómeno cuyo carácter expresivo no es arbitrario, sino que está determinado por la propia fisiología del individuo que lo experimenta. Resumiendo, los conocimientos que aporta actualmente la psicología de las emociones nos permitían establecer una serie de estados emocionales fundamentales o básicos y nos indicaban que cada una de ellos está asociado a un tipo de conciencia interna, de conducta expresiva y de modificaciones fisiológicas. En consecuencia, si conseguíamos observar en los actores informantes alguno de los indicadores fisiológicos que se alteran con la aparición de las distintas emociones, también podríamos saber objetivamente si el actor había experimentado realmente alguna emoción mientras interpretaba cada uno de los discursos.

Nuestro corpus fue construido a partir de las locuciones de 8 actores, 4 hombres y 4 mujeres. Cada uno de ellos tenía que interpretar dos textos distintos en diferentes estados emocionales y con tres distintos grados de intensidad en cada estado emocional. Concretamente, en cada una de las lecturas emocionadas tenía que quedar reflejada una de las siguientes emociones: *sorpresa*, *alegría*, *deseo*, *rabia*, *tristeza*, *asco* y *miedo*. El objetivo final era que el corpus contuviera cada uno de esos estados emocionales expresados con suficiente naturalidad y verosimilitud, para que luego fuese posible analizarlos y modelizarlos acústicamente.

Para conseguir todo esto se solicitaba al actor que interpretase cada emoción, por ejemplo la *sorpresa*, en primer lugar con poca intensidad emocional, luego con una intensidad emocional media y, finalmente, con una intensidad emocional alta. Con objeto de facilitar al actor su trabajo de puesta en situación, una vez aclarada la emoción que tenía que interpretar se le proponía una frase concreta para sugerirle cada grado de intensidad emotiva. Así, en el caso de la *sorpresa* las frases de sugerencia fueron las siguientes: para proponer poca intensidad se le dijo al actor "*algo te extraña*", para la intensidad media se le dijo "*has quedado pasmado*", y para sugerirle una intensidad emocional muy fuerte se le propuso "*estás completamente desconcertado*". Lógicamente, una vez planteada cada situación dramática se iniciaba una grabación sonora, con lo que todas las interpretaciones quedaron registradas en cinta magnetofónica. Este proceso se repetía con todas las emociones, proponiendo siempre al actor frases de sugerencia distintas y adecuadas a cada emoción concreta. Por ejemplo, en el caso de la *alegría* las frases de sugerencia fueron: "*te sientes contento*", "*estas entusiasmado*" y "*te has puesto completamente eufórico*". Posteriormente, se repetía de nuevo todo este ciclo interpretativo con un segundo texto, de modo que el corpus quedaría compuesto finalmente por las interpretaciones de 2 textos, por parte de 8 actores, interpretando cada texto con 7 emociones básicas y en tres grados de intensidad distintos cada una de ellas. Es decir, dispondríamos de un corpus con $2 \times 8 \times 7 \times 3 = 336$ interpretaciones emocionadas distintas.

Probablemente, los actores consiguieron emocionarse realmente sólo en algunas de estas 336 interpretaciones, el resto eran interpretaciones puramente "técnicas", es decir, fingidas. Debíamos, pues, diseñar un método que nos permitiese decidir de forma objetiva qué interpretaciones estaban realmente vinculadas a alguna emoción y qué otras no lo estaban; de esa manera, una vez desarrollado el

análisis acústico, sería posible vincular con garantías metodológicas suficientes ciertos parámetros acústicos de la voz con estados emocionales concretos.

Para conseguir ese control durante las interpretaciones, basándonos en los conocimientos sobre psicología de las emociones expuestos más arriba, decidimos observar durante cada interpretación del actor algunas constantes fisiológicas que fuesen capaces de revelarnos objetivamente su estado emocional, concretamente se decidió estudiar el ritmo cardiaco y la presión arterial. Los datos sobre la presión arterial y ritmo cardiaco de cada actor se obtuvieron utilizando tensiómetros "Omron R1" modelo HEM-601 R1 distribuido por la empresa *Omron Corporation*. La toma de la presión y pulso se realizaba inmediatamente después de que el actor finalizara cada una de sus interpretaciones. Se realizaron, en total, 42 tomas de la presión arterial y el ritmo cardiaco para cada uno de los 8 actores. Cada una de las mediciones aportaba tres datos: presión arterial sistólica en mmHg, presión arterial diastólica en mmHg y ritmo cardiaco en número de latidos por minuto.

Pudimos observar que, efectivamente, se manifestaban unas tendencias claras en varias emociones:

–ALEGRIA: indujo a subir la presión sistólica en 5 de los 8 actores.

–ASCO: indujo a bajar la presión sistólica en 5 de los 8 actores.

–MIEDO: indujo a bajar la presión sistólica en 7 de los 8 actores.

–SORPRESA: indujo a subir la presión sistólica y presión diastólica en 6 de los 8 actores.

–TRISTEZA: indujo a bajar el ritmo cardiaco 5 de los 8 actores.

–No se observó ninguna tendencia clara para el DESEO ni para la RABIA.

Observamos, también, que frente a una emoción se producían dos respuestas posibles: 1) el locutor se excitaba y tendía a la agitación, lo cual provocaba una tendencia a subir su ritmo cardiaco y bajar su presión arterial; 2) el locutor se inhibía y tendía a la depresión, entonces tendía a desencadenarse la tendencia inversa, es decir, a bajarle el ritmo cardiaco y subirle la presión arterial.

Desarrollo de tests de recepción con todos los mensajes del corpus

De todos modos, el control fisiológico durante la interpretación de los locutores no nos garantizaba, en principio, que apareciese una asociación sistemática entre los momentos de "emoción real" y las interpretaciones "óptimas". Lógicamente, un individuo puede estar emocionado y reprimir la expresión de esa emoción; o bien, puede no estarlo pero fingir expresivamente que lo está, consiguiendo, además, engañar al receptor. No podíamos olvidar que un actor está entrenado para fingir, y que nuestro objeto de estudio no era la emoción real sino la capacidad para expresar esa emoción a través de la voz.

La objetivación definitiva de nuestro corpus se realizó, pues, sometiendo cada una de las 336 interpretaciones emocionadas al juicio de más de 60 sujetos experimentales. Organizados en distintos grupos, se hizo escuchar a los sujetos experimentales series concretas de interpretaciones. Y tras la audición de cada voz, se solicitaba mediante un test a cada oyente que respondiese a las tres cuestiones siguientes: 1) debía concretar qué emoción o emociones reconocía en la 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 con qué grado. El corpus fue expuesto, en total, a más de un millar de sujetos experimentales.

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. Y permitió, además, vincular las emociones que fueron reconocidas en cada voz, con las distintas tendencias observadas en las variaciones fisiológicas. En esta última fase de validación, comprobamos, efectivamente, que los oyentes consideraban bastante más

verosímiles las interpretaciones que tenían asociados unos datos fisiológicos muy claramente desviados de sus promedios; y pudo observarse, también, que este fenómeno se producía con independencia de la emoción. No obstante, esta vinculación sistemática entre la desviación extrema de los datos fisiológicos con una mayor verosimilitud sólo se cumplía en algunos locutores, en cambio en otros no fue así. Probablemente esto se debe a la distinta capacidad de los actores que actuaron como informantes de nuestro corpus, para autoinducir su estado psicológico durante las interpretaciones en un estado emocional objetivo y real.

Análisis estadístico para la contrastación de las hipótesis generales y la selección de los mensajes comunicativamente “óptimos”

Los tests de recepción nos tenían que permitir la contrastación definitiva de la hipótesis general de nuestra investigación y decidir cuáles eran las interpretaciones comunicativamente óptimas y, por tanto, idóneas para ser analizadas.

El análisis estadístico de los tests de recepción demostró, efectivamente, que un locutor interpretando el mismo texto puede expresar sin problemas emociones radicalmente distintas. Y que estas emociones son perfectamente reconocibles de manera coherente y unívoca por un porcentaje de sujetos absolutamente superior al que correspondería al azar. Es decir, la hipótesis general fue ampliamente contrastada para las 7 emociones básicas con las que estuvimos trabajando.

La conclusión más directa de los resultados obtenidos en esta etapa de la investigación era que, efectivamente, un orador expresa emociones codificadas en el sonido del habla, y que esta expresión es completamente independiente del texto utilizado. Y puesto que estas emociones habían sido reconocidas e identificadas por los sujetos experimentales sin depender de los contenidos textuales, eso significaba, también, que el sonido de los discursos del corpus contenía lo que estábamos buscando. Es decir, que las distintas alteraciones sonoras del habla asociadas a cada una de las 7 emociones podían ser localizadas y formalizadas estudiando los mensajes de nuestro corpus.

El paso siguiente de esta fase era seleccionar de entre los 336 mensajes del corpus unos cuantos que pudiésemos considerar como “comunicativamente óptimos”. Cuando hablamos aquí de interpretaciones “óptimas” nos estamos refiriendo a aquellas voces que fueron asociadas coherentemente a una emoción concreta por un porcentaje superior al 70% de los sujetos experimentales que las juzgaron y que, además, fueron valoradas por esos mismos sujetos como interpretaciones de una verosimilitud superior a la media. Aplicando este criterio localizamos 4 ó 5 para cada emoción que cumplían esas características, con una única excepción: el *asco*.

Aunque al aplicar las pruebas de hipótesis observamos que el *asco* también había sido identificado con unos porcentajes por encima de los que corresponderían al azar; en realidad, solamente una de las interpretaciones había conseguido un grado de identificación de esa emoción por encima del 50%. Dicho de otro modo, podíamos afirmar que el *asco* era identificable, pero no se manifestaba aisladamente con la suficiente claridad, sino que esta emoción aparecía sistemáticamente vinculada a la *rabia* o al *miedo*. Podíamos disponer pues de interpretaciones con *asco* muy bien calificadas por los sujetos experimentales, pero siempre con alguna otra emoción asociada. Esto nos llevo, finalmente, a optar, también, por añadir a la selección algunas interpretaciones con dos emociones que habían quedado muy claramente asociadas entre sí durante el análisis estadístico y a las que hemos denominado “híbridas”. El resultado último fue una selección de 5 interpretaciones de la *alegría*, 4 del *deseo*, 1 del *asco*, 4 del *miedo*, 5 de la *rabia*, 4 de la *sorpresa* y 5 de la *tristeza* como emociones puras además de 2 de “*alegría + sorpresa*”, 1 de “*deseo + rabia*”, 3 de “*deseo + miedo*”, 3 de “*asco + rabia*”, 1 de “*asco + tristeza*”, 3 de “*tristeza + miedo*”, 1 de “*miedo + sorpresa*”, 1 de “*miedo + rabia*”, 1 de “*sorpresa + alegría*” y 1 de “*deseo + sorpresa*”, como emociones híbridas.

Análisis instrumental (acústico) de los mensajes “óptimos”, obtención de datos y extracción de modelos formales

Con los mensajes que vamos a estudiar ya seleccionados y teniendo la garantía científica de que contienen aquello que buscamos, la etapa siguiente era desarrollar un análisis de las interpretaciones elegidas que fuese capaz de sacar a la luz las características sonoras específicas de cada emoción.

Esta etapa del estudio requería una especialización por parte del grupo de investigadores que ha de desarrollar un proyecto de este tipo. Es cierto que hablar de la localización de estructuras formales utilizando instrumentos de medida supone un planteamiento genérico y abierto, que es aplicable a numerosos objetos de estudio. No obstante, desde la perspectiva metodológica que estamos proponiendo³, precisábamos de una serie de aportaciones de disciplinas contiguas. En concreto, necesitábamos instrumentos de análisis y medición acústica, y de una serie de conocimientos previos que podían aportar la lingüística general, la prosodia y la fonética experimental. Necesitábamos, también, saber, a priori, cuáles eran las estructuras sonoras vinculadas con el habla neutra (no emocionada), y qué tipo de procesos se siguen en estas disciplinas para encontrar modelos formales en el habla con objeto de adaptarlos a nuestras propias necesidades.

De hecho, en el “*Laboratorio de la Voz y el Sonido*” del Dpto. de Comunicación Audiovisual y Publicidad, disponíamos ya de buena parte de la infraestructura necesaria para esta investigación, y esa infraestructura pudo ser complementada, además, por la ayuda económica aportada por la DGICYT. Por otra parte, la propia trayectoria de trabajo del grupo que presenta este estudio cubría esa necesidad de especialización que mencionábamos más arriba (Rodríguez, 1984, 1988, 1989, 1992 y 1998; Martínez, 1994; Montoya, 1996, 1998; Lázaro, 1998).

En el momento en que se escriben estas líneas estamos todavía trabajando en la fase de análisis instrumental. En consecuencia, aún no disponemos de modelos definidos que puedan ser presentados como resultado de esta etapa de la investigación. No obstante, expondremos los criterios generales con los que estamos trabajando y los aspectos principales en que estamos centrando nuestra atención.

Nuestra aproximación al sonido de las emociones desarrolla el análisis acústico desde una perspectiva suprasegmental, es decir, estudiando cómo evoluciona el habla a lo largo del tiempo. Entendemos, además, que la expresividad sonora de la voz debe abordarse desde una perspectiva múltiple, que sea capaz de observar la evolución sincrónica de distintos parámetros acústicos y las relaciones entre ellos.

En coherencia con estos criterios, el trabajo de análisis acústico que ya hemos realizado tiende a romper con la idea, hasta ahora dominante, de que la expresión emocional está asociada esencialmente a las variaciones tonales. Hemos observado, por ejemplo, que mientras las emociones positivas parecen tender a aumentar simultáneamente la variabilidad del contorno tonal, a mantener homogénea la intensidad y a disminuir los tiempos de pausa; las emociones negativas parecen comportar a la vez una estabilización del tono, el aumento de las variaciones de la intensidad y el incremento de la duración de las pausas. Nuestras observaciones parecen mostrar, también una relación clara entre la *alegría* con un aumento de intensidad en las frecuencias graves del espectro de la voz. En cambio, la *rabia* se muestra vinculada regularmente con la disminución de los graves combinada con un aumento de intensidad en las resonancias agudas.

En suma, trabajamos desde una perspectiva múltiple que intenta contemplar, a la vez, la entonación, la evolución de la intensidad, la composición espectral y la estructura rítmica como un sistema de variables que interactúan sincronizadas en el tiempo.

Esta última etapa de la metodología consiste en estudiar las correlaciones que existen entre los datos que se obtuvieron en las pruebas de recepción, con los datos que se conseguirán en el análisis acústico. El estudio de las correlaciones puede realizarse siguiendo una estrategia "masiva", o bien siguiendo una estrategia "selectiva". En el primer caso, desarrollaríamos el análisis instrumental extensivamente sobre todo el corpus, para unir, luego, en un sólo fichero los datos obtenidos en esta medición con los conseguidos antes en el estudio de recepción. Y, a partir de ese momento, se observarían conjuntamente para buscar las correlaciones que existen entre ambos mediante procedimientos estadísticos. En el segundo caso, es decir, si seguimos la estrategia "selectiva", el análisis instrumental sólo se realizará sobre un número reducido de mensajes del corpus que habrán sido antes minuciosamente seleccionados estudiando sus efectos sobre los receptores.

En la investigación que exponemos se optó por la estrategia *selectiva* por dos razones: en primer, lugar porque la propia extensión del corpus (336 interpretaciones emocionadas) hubiese hecho la tarea de las mediciones prácticamente inabarcable. No obstante, la razón segunda y fundamental tuvo un carácter estrictamente científico: puesto que disponíamos ya de experiencia previa tanto en la utilización de esta metodología como en el estudio de la expresividad sonora del habla, el planteamiento del análisis acústico podía ser mucho más preciso y concreto. Sabíamos lo que buscábamos y sabíamos dónde buscarlo. En consecuencia, no era necesario realizar una obtención masiva de datos. Nuestro conocimiento previo del objeto de estudio nos permitía saltarnos la fase epistemológica inicial que ha de seguir toda investigación empírica para encontrar las primeras tendencias objetivas de un fenómeno, y que sólo suele ser resoluble con una observación extensiva apoyada en la estadística.

En definitiva, pues, la estrategia que decidimos seguir en esta etapa final fue la siguiente. Partíamos de un minucioso trabajo de selección y clasificación de las voces del corpus, por tanto, si el análisis instrumental revelaba, por ejemplo, que determinada forma acústica estaba presente en todas las interpretaciones de la *rabia* sin aparecer en ninguna otra más, entonces ya sería posible inducir de eso que el carácter sonoro localizado es específico de la *rabia*. Dicho de otra modo: la información previa que habíamos obtenido sobre las voces que íbamos a analizar nos permitiría inducir los modelos acústico-emocionales que estábamos buscando, sin necesidad de proceder a un análisis acústico del corpus al completo.

CONCLUSIONES

En tanto que la última parte de la investigación que presentamos está todavía sin concluir, aun no podemos exponer nuestras conclusiones sobre los modelos de relación acústico-emocional en el habla. No obstante, consideramos que tanto los resultados que hemos obtenido ya en el estudio que nos ocupa, como los conseguidos en otras cuatro investigaciones que fueron desarrolladas siguiendo este mismo método con objetos de estudio diversos (Rodríguez, 1989, 1995; Montoya, 1996; Tena, 1998), nos permiten presentar unas conclusiones sólidas sobre los aspectos de nuestro trabajo que configuran un nuevo instrumento metodológico.

A continuación expondremos, pues, las cuestiones que consideramos esenciales en el método de trabajo que proponemos y que más arriba nombrábamos como *metodología de análisis instrumental*.

Descripción global del método

El método de *análisis instrumental* se desarrolla en dos grandes bloques complementarios:

1. El análisis objetivo de las formas sonoras y/o visuales de los mensajes utilizando instrumentos de medida (ya existentes, o desarrollados *ad hoc*).
2. El estudio individualizado de los efectos que ha producido cada uno de esos mensajes concretos sobre un grupo suficientemente amplio de receptores, utilizando instrumentos de control objetivo de la percepción.

La búsqueda posterior de las relaciones entre estos dos grandes bloques de trabajo, es decir, entre las formas sonoras o visuales localizadas y los efectos que éstas han producido en la recepción, es lo que nos dará respuestas concretas sobre el funcionamiento de los mecanismos expresivos audiovisuales.

Su aplicación

La *metodología de análisis instrumental* se adecua y da respuestas eficaces fundamentalmente en las investigaciones relacionadas con el funcionamiento de los lenguajes. Es un método adecuado para resolver, por ejemplo, preguntas como las siguientes:

- ¿Qué tipo de uso de los colores es necesario hacer para aumentar el efecto persuasivo de los mensajes?
- ¿Cómo afecta la escenografía de los informativos de TV a la credibilidad de la información?
- ¿Cómo debe utilizar su voz un locutor para conseguir un mayor índice de atención y memorización de sus discursos?
- ¿Por qué influyen la situación y el tratamiento visual de un bloque de texto sobre una página en su elección para la lectura y cómo podemos priorizar su elección?

Dicho de otro modo, es un método especialmente adecuado para las investigaciones que persiguen conocer el funcionamiento de los procesos expresivos y narrativos en los discursos audiovisuales, o que intentan saber cómo se puede mejorar su rendimiento comunicativo.

Sus exigencias

Una cuestión importante a tener en cuenta es que este método exige especialización en el objeto de estudio con el que se trabaja. Exige, también, conocimientos amplios en disciplinas contiguas a la comunicación a fin de poder utilizar o, en su caso, desarrollar los instrumentos de medición y de control necesarios. De hecho, si trabajamos desde esta perspectiva, la comunicación audiovisual plantea ya desde un primer momento dos grandes objetos de estudio: 1) la *luz* y las *formas visuales*, 2) el *sonido* y las *formas auditivas*, que exigen formaciones previas específicas muy diferentes. Mientras el primer objeto exige un conocimiento profundo de la psicología de la percepción visual, la física y los instrumentos de medición relacionados con la luz; el segundo lo exige de la percepción auditiva, de la física acústica y de toda la instrumentación relacionada con el sonido.

Es imprescindible contemplar a la vez mensaje y receptor

En la concepción de esta metodología es fundamental observar el sistema *MENSAJE-RECEPTOR* como una globalidad no disociable. Es cierto que es en el mensaje donde están codificadas físicamente las formas auditivas y visuales que va a utilizar quien reciba el mensaje, pero es imprescindible considerar que todo mensaje contiene muchas más características acústicas y lumínicas que las que van a procesar y utilizar los receptores. Y que solamente el receptor puede

darnos las pautas sobre cuáles son las partes realmente útiles en toda la complejidad física de un mensaje audiovisual.

Cuando sólo estudiamos el mensaje, nuestro punto de vista se centra exclusivamente en el emisor, ignorando cuál es la influencia real de los mensajes en el público que los recibe. Y cuando solamente estudiamos al receptor, en realidad, estamos cerrando los ojos a los estímulos que causan los resultados de una comunicación.

El proceso comunicativo como objeto de estudio

La *metodología de análisis instrumental* establece un carácter diferencial muy claro respecto a los estudios sobre comunicación apoyados en métodos empíricos de base sociológica. Mientras los estudios sobre recepción y efectos que siguen los modelos metodológicos más clásicos toman como objeto de estudio al receptor, el método que proponemos toma como objeto el proceso de la comunicación. El *método de análisis instrumental* no estudia los efectos sobre los sujetos, sino cómo se relacionan los mensajes con los receptores.

Pero ¿cuál es la forma de concretar este matiz conceptual aparentemente tan peregrina y sutil?

La forma de desplazar el punto de vista desde el receptor hasta el fenómeno comunicativo como sistema global consiste, esencialmente, en estudiar los datos resultantes de las investigaciones sobre recepción tomando como referencia las variables vinculadas a las características del mensaje, en lugar de las vinculadas a las de los oyentes. Es decir, no son las especificidades de la muestra de receptores las que hemos de tomar como variables independientes para observar su influencia en los resultados, sino las formas y los modelos de los discursos utilizados. Habitualmente se analiza cómo influye la edad, el sexo, la formación o la clase social en los efectos del mensaje; ahora, lo que hemos de tomar como referencia es cómo actúan sobre los efectos del mensaje el tono de la voz, la velocidad de locución o la duración de las pausas.

De contar a medir

También nos parece una cuestión importante el salto conceptual que existe entre una cuantificación que observa la cantidad de elementos de la clase "A" que cumplen, o no, determinada característica; y una cuantificación que asigna cifras a las diferencias concretas que observamos entre esos mismos elementos de la clase "A". Hablamos, obviamente, de las diferencias entre *contar* y *medir*.

Mientras para el primer caso (contar) basta con definir con suficiente precisión alguna propiedad de los elementos del objeto que estamos estudiando; el segundo (medir) exige encontrar o diseñar una magnitud, es decir, una escala genérica de valores numéricos definida respecto a un patrón de referencia estable, de modo que sea posible desarrollar comparaciones objetivas entre los elementos del objeto estudiado. En tanto que la *metodología de análisis instrumental* propone que puedan desarrollarse comparaciones objetivas dentro de las distintas categorías que estamos estudiando, estamos precisando de un salto epistemológico esencial. Necesitamos pasar del estadio de contar al de medir.

Para estudiar, por ejemplo, la relación entre el tono de voz y la sensación de *rabia* no nos basta con contar qué proporción de interpretaciones con tono "agudo", con tono "medio" o con tono "grave" generan sensación de *rabia* en nuestros sujetos experimentales. Todas las mujeres tienen la voz mucho más "aguda" que los hombres; a pesar de ello, entre las interpretaciones femeninas tanto encontraremos voces que producirán sensación de *rabia* como otras que no la producirán; y lo mismo ocurrirá con las voces "graves" masculinas. Es evidente, entonces, que contar y establecer proporciones entre las voces con la propiedad "*emitir un tono de voz agudo*", la de "*emitir un tono de voz medio*" o la de "*emitir*

un tono de voz grave” resulta completamente insuficiente. Es necesario profundizar mucho más en el conocimiento de esa variable y encontrar una magnitud escalar que nos permita establecer comparaciones objetivas mucho más precisas entre los distintos niveles tonales de las voces. O lo que es lo mismo, es necesario pasar de contar elementos a medir parámetros.

En suma, en tanto que el método necesita analizar las variables físicas vinculadas a la visión y la audición para descubrir modelos formales con valor expresivo, exige, también, un conocimiento profundo de estas variables. Es decir, si utilizamos el *método de análisis instrumental* no nos bastará con contar, sino que deberemos encontrar o desarrollar instrumentos para medir.

El conocimiento de la percepción como eje central

Si la gama de frecuencias sonoras que percibe cualquier ser humano es prácticamente idéntica a la que perciben todos los demás y la gama de frecuencias lumínicas también lo es, si el sistema receptor de información externa responde a las mismas necesidades biológicas, espaciales, cinéticas, alimenticias, táctiles, de temperatura, etc., probablemente no incurriremos en ningún error afirmando que es posible localizar variables universales en los mecanismos humanos de interpretación de los mensajes sonoros y visuales. De acuerdo con esta reflexión, esta metodología toma como punto de partida la percepción del ser humano y se desarrolla a partir de ese criterio básico central.

Notas

¹ Estudio desarrollado en el marco del proyecto “*Modelización acústica de la expresión emocional en el español*”, financiado por la DGICYT (PB94-0730).

² Como habrá observado el lector, de forma paralela a la exposición de nuestro estudio sobre el habla estamos proponiendo, también, nuevas herramientas metodológicas que aportan una perspectiva completamente distinta a la orientación ya clásica de los paradigmas sobre recepción y efectos de la comunicación (Wolf, 1987, 1994).

³ El lector puede encontrar en la obra *La dimensión sonora del lenguaje audiovisual* (Rodríguez, 1998, pp. 19-32) una exposición mucho más amplia y detallada en torno a estos planteamientos metodológicos.

Referencias

- BALSEBRE, A. (1994). *El lenguaje radiofónico*. Madrid: Cátedra .
- CEBRIAN , M. (1994). *Información radiofónica. Mediación técnica, tratamiento y programación*. Madrid: Síntesis.
- DANTZER, R. (1989). *Las emociones*. Barcelona: Paidós.
- HILLS, G. (1981). *Los informativos en radiotelevisión*. Madrid: Instituto Oficial de Radio y Televisión.
- IZARD, C. (1982). *Measuring emotions in infants and children*. Cambridge: Cambridge University Press.
- LÁZARO, P. (1998). Tecnología del audio: investigación y didáctica de la expresión oral. *Actas de II Congreso Internacional de Formación y Medios*. Segovia, julio.
- MARTÍNEZ, G. (1994). *Persuasión publicitaria. La voz del personaje y la implicación emotiva*. Tesis de licenciatura. Barcelona: Dpto. de Comunicación Audiovisual y Publicidad de Universidad Autónoma de Barcelona.
- MCLEISH, R. (1985). *Técnicas de creación y realización en radio*. Madrid: Instituto Oficial de Radio y Televisión.

- MERAYO, A. (1992). *Para entender la radio. Estructura del proceso informativo radiofónico*. Salamanca: Universidad Pontificia de Salamanca.
- MONTOYA, N. (1996). *El papel de la voz en la publicidad audiovisual dirigida a los niños*. Tesis de licenciatura. Barcelona: Dpto. de Comunicación Audiovisual y Publicidad de la Universidad Autónoma de Barcelona.
- MONTOYA, N. (1998). La persuasión publicitaria a través de la voz y su aplicación en la educación. *Actas del II Congreso Internacional de Formación y Medios*. Segovia, julio.
- MUÑOZ, J. J. y GIL, C. (1986). *La radio. Teoría y práctica*. Madrid: Instituto Oficial de Radio y Televisión.
- REEVE, J. M. (1994). *Motivación y emoción*. Madrid: McGraw-Hill/Interamericana de España.
- RODRÍGUEZ, A. (1984). *La voz en la radio. Manipulaciones y técnicas de expresión*. Tesis de licenciatura. Barcelona: Dpto. de Comunicación Audiovisual y Publicidad de la Universidad Autónoma de Barcelona.
- RODRÍGUEZ, A. (1988). Máquinas que hablan y escuchan. *Telos*, 16, 117-125.
- RODRÍGUEZ, A. (1989). *La construcción de una voz radiofónica*. Tesis doctoral. Barcelona: Dpto. de Comunicación Audiovisual y Publicidad de la Universidad Autónoma de Barcelona.
- RODRÍGUEZ, A. (1992). Crítica del paradigma fonético como base del reconocimiento de formas en el tratamiento automático del habla. En *Actas de las Jornadas Nacionales de Acústica* (pp. 141-144). Pamplona: Sociedad Española de Acústica.
- RODRÍGUEZ, A. (1995). Una nueva propuesta metodológica en torno al ritmo visual: aplicación del método de análisis instrumental al ritmo visual de una telenovela y un telefilme norteamericano. *Análisi*, 18, 87-107.
- RODRÍGUEZ, A. (1998). *La dimensión sonora del lenguaje audiovisual*. Barcelona: Paidós.
- TENA, D. (1998). *La influencia de la composición gráfica en la elección de un bloque de texto escrito*. Tesis Doctoral. Barcelona: Dpto. de Comunicación Audiovisual y Publicidad de la Universidad Autónoma de Barcelona.
- WOLF, M. (1987). *La investigación de la comunicación de masas*. Barcelona: Paidós.
- WOLF, M. (1994). *Los efectos sociales de los media*. Barcelona: Paidós.

Sobre los autores

Ángel Rodríguez Bravo es doctor en Ciencias de la Información y profesor titular de "Teoría de la expresión acústica". Premio Extraordinario de Doctorado en 1991 y IV Premio a la Investigación sobre Comunicación de Masas de la Generalitat de Cataluña. Su principal trabajo publicado es "*La dimensión sonora del lenguaje audiovisual*" (Paidós, 1998). En la actualidad dirige el *Laboratorio de la Voz y el Sonido*, del "Institut de la Comunicació".

Patricia Lázaro Pernias, Norminanda Montoya Vilar, Josep María Blanco Pont, Dolors Bernadas Suñé, J. Manel Oliver Comes y Ludovico Longhi son miembros del equipo de investigación.

Dirección de los autores:
 Universidad Autónoma de Barcelona.
 Dpto. de Comunicación Audiovisual y Publicidad.
 Edificio I
 08193 Bellaterra (Barcelona) Spain.
 Tel.: 93-5811643
 Fax.: 93-5812005, 93-5811540
 E-Mail: arbravo@intercom.es