

MÚSICA Y AUDICIÓN EN LOS GÉNEROS AUDIOVISUALES

Josep Gustems (coord.)

>> 13



COMUNICACIÓN
ACTIVA
AUDIOVISUAL



Índice

Prólogo	9
Antoni Mercader, UB	
Presentación.....	11
Josep Gustems, UB	
MODOS DE AUDICIÓ N EN LOS AUDIOVISUALES	
El análisis multimodal en la escucha de los audiovisuales	15
Josep Gustems, UB, Caterina Calderón, UB	
Las Unidades Semióticas Temporales (UST), estrategia perceptiva y vía analítica para la música	29
Antonio Alcázar, UCLM	
Los modos de escucha televisiva.....	53
M. Amparo Porta, UJI	
Cine para los oídos: el lenguaje del sonido grabado	69
Josep Maria Mayol, UB	
LA MÚSICA EN LOS GÉNEROS AUDIOVISUALES	
Estructuras del clip musical.....	89
Joan Marimón, ESCAC	

Del aria al primer plano. Apuntes para el estudio de la ópera en el cine.....	103
Jaume Radigales, UB, URL	
La música de los dibujos animados.....	115
Jaume Duran, UB	
La música en los videojuegos.....	131
M. Melania Gómez, UB	
Música y publicidad audiovisual.....	143
Josep Gustems, UB, Caterina Calderón, UB	
La dimensión sonora de los tráileres.....	155
Francesc Llinares, UB	
Música y sonido en el ‘lipdub’.....	169
Diego Calderón, UB	

/ El análisis multimodal en la escucha de los audiovisuales

Josep Gustems Carnicer, *Universidad de Barcelona*
Caterina Calderón Garrido, *Universidad de Barcelona*

> Introducción

Los audiovisuales son productos complejos, resultado de la intervención de distintos creadores, técnicos y artistas con roles, formación y puntos de vista completamente diferentes. Para entender mejor el alcance de cada uno de ellos en la creación audiovisual y, por ende, en la formación de los estudiantes de Comunicación Audiovisual, es preciso dedicar un espacio al estudio de la multimodalidad como forma de interacción, de trabajo colaborativo y de marco teórico.

El análisis multimodal surge gracias al paradigma interpretativo y al incremento de abordajes multidisciplinares en las investigaciones, y trae aparejadas nuevas necesidades metodológicas, como el paso de una postura *realista* a una *relativista*, sobre todo en los audiovisuales etnográficos. Este cambio desempeña un rol activo en la construcción del conocimiento (D'Angelo et al., 2009), y da forma a los significados mediante los recursos disponibles para los usuarios, quienes son activos y no solo reproducen significados, sino que los crean (Manghi, 2009).

En el ámbito de la lingüística y de las ciencias sociales, una de las manifestaciones de ese cambio es el auge por los estudios multimodales, que puso en evidencia la necesidad de contar con un proceso de transcripción de la realidad adecuado, que facilitase la descripción y el posterior pasaje al análisis y a la interpretación de los discursos multimodales. En el terreno académico, el análisis multimodal se ha llevado a cabo siguiendo las indicaciones de ilustres lingüistas como G. Kress (U. Londres), C. Thurtow (U. Washington) o N. Fairclough (U. Lancaster).

Modos de audición en los audiovisuales

Un discurso es multimodal porque está integrado por más de una forma de expresión semiótica (D'Angelo et al., 2009). El texto escrito, en muchos casos, ya no es el centro, sino que funciona como complemento para potenciar el mensaje creado en una red interrelacionada de lenguajes que le dan significado (Kress y Van Leeuwen, 2001). Kress (2001) sostenía que la multimodalidad de alguna manera estudia una obviedad porque no hay textos monomodales. En otras palabras, no hay textos sin materialidad, no hay textos orales sin voces, no hay textos escritos sin soporte material.

La multimodalidad es un posicionamiento ecléctico de carácter más técnico que teórico. Existe cierta política respecto a la multimodalidad y a las convenciones sociales y académicas respecto al grado de importancia asignado a los diferentes modos de comunicación y a sus combinaciones. El uso o preeminencia de cada modo depende en cierta manera del grado de ambigüedad que permite y, por tanto, del control que puede establecerse en cada uno de los medios. Las combinaciones y yuxtaposiciones multimodales responden a la intención del poder (Lemke, 2002).

> En busca de una definición de multimodalidad

El análisis multimodal del discurso es el estudio de la intersección e interdependencia de varias modalidades de comunicación en un contexto dado, a modo de puente entre modos de significación (O'Halloran, 2008). La comunicación sucede a través de más de un modo de comunicación y, en consecuencia, es inherentemente multimodal.

En general, los sistemas multimodales benefician a los usuarios, puesto que tales modalidades durante una interacción deben incrementar su ejecución completando una tarea (Sulaiman et al., 2008). La multimodalidad se ocupa de todas las maneras que tenemos de construir significado —los modos de representación (Kress y Van Leeuwen, 2001).

La Teoría Multimodal es una perspectiva de Análisis Sistémico-Funcional (SF-MDA) (Halliday, 1978), que postula que la comunicación sucede a través de diferentes canales o modos de significación de forma simultánea (texto, imágenes, gráficos, diseño, sonido, música, gesto, habla, banda sonora, 3D, etc.). Así, para entender el significado de un acto comunicativo, debemos tener en cuenta todos los modos de significación que se emplean en él. Por ello, estos modos de significación no pueden ser estudiados de mane-

ra aislada, ya que están estrechamente conectados entre sí (Pujolá y Montmany, 2010).

Los sistemas de diálogo multimodal (*Multimodal Dialogue Systems*, MDSs), entre los que se cuentan los audiovisuales y el *software*, utilizan varios canales de comunicación (p. ej.: habla, texto, expresiones faciales, gestos, etc.) mediante micrófonos y altavoces, cámaras para visión artificial, guantes de datos, pantallas sensibles al tacto, etc., proporcionando una interacción más flexible, cómoda y adaptativa, la cual permite reducir los errores en la comunicación (López-Cozar, Callejas y Gea, 2005).

> Bases psicobiológicas de la multimodalidad

La información combinada aumenta los índices perceptivos de los receptores (López-Cozar, Callejas y Gea, 2005). El mensaje multimodal permite una interpretación más fiable de los mensajes e intenciones comunicativas. Según Lemke (1998), el resultado de la semiosis multimodal es multiplicativo. El potencial de significado, la capacidad de recursos significativos y las construcciones multimodales son el producto lógico, en un sentido multiplicativo, de las disposiciones de cada uno de los sistemas de recursos semióticos constituyentes (Lemke, 2002).

El modo de procesar la información humana se puede diferenciar en: presentación multimedia, memoria sensorial, memoria de trabajo y memoria a largo plazo. La Teoría del «Doble Canal» sugiere que los estudiantes aprenden mejor cuando se procesan varios canales a la vez. Los resultados experimentales mostraron estadísticamente mejores resultados de aprendizaje para el grupo experimental en los elementos musicales. También se hallaron efectos positivos para la motivación y las actitudes hacia la enseñanza para el alumno del grupo multimodal (Yu et al., 2010).

El apoyo para tales nociones de percepción intermodal se suministra mediante la convergencia de pruebas neurológicas, lo que sugiere que áreas subcorticales específicas (como el colículo superior) combinan información visual, auditiva y somatosensorial en una representación espacial amodal (Spence y Driver, 1997).

Algunos estudios anatómicos comparativos también apoyan la idea de que una de las funciones primarias de la localización del sonido puede ser la de controlar la orientación de las manifestaciones visuales (Spence y Driver, 1997), lo que

sugiere un cierto diseño multimodal en las funciones perceptivas. Por ejemplo, Wundt calculó umbrales de simultaneidad en percepción intramodal: 2 ms para sonido, 27 ms para tacto y 43 ms para vista. En intermodalidad (audio respecto a vídeo), la asincronía se da más allá de los -41 ms y $+45$ ms (Levitin et al., 2000).

Asimismo, otros estudios evidencian la activación de áreas del cerebro asociadas con el procesamiento visoespacial durante las tareas relacionadas con la música, lo que demuestra que existe cierto entrecruzamiento biológico en los modos perceptivos y comunicativos (Zatorre, Evans y Meyer, 1994).

> Modos de comunicación y metafunciones

Cada uno de los sistemas semióticos posee una carga o potencial comunicativo diferente, las llamadas *affordances*, que corresponden a lo que es posible significar con cada modalidad semiótica (Kress y Van Leeuwen, 2001). Aun teniendo características comunes, como la dimensión espacio-temporal de algunos modos, los potenciales de cada modo son diferentes (Manghi, 2009) y, por tanto, se escogen o se explotan según el efecto que se quiera conseguir, dependiendo de la intención comunicativa (Pujolá y Montmany, 2010). El mensaje es el lugar de encuentro, es decir, la intersección de diferentes códigos interdependientes en una red interrelacionada de lenguajes que le otorgan significado (*orquestación semiótica*). Cada modo, en un texto multimodal, solo lleva un significado parcial (Manghi, 2009).

Las distinciones que caracterizan cada modo son notables. Norris (2004) pone el acento en el cuerpo: para él, existen unos modos de comunicación corporales (*embodied*) y otros no corporales (*disembodied*). En cambio, otros investigadores ponen el énfasis en las características y significados del contexto: Baldry y Thibault (2006) usan los conceptos de *contexto situacional* y *contexto cultural* para delinear distinciones importantes entre modalidades.

Respecto a las formas de significación de cada modo, el lenguaje representaría significados referidos a tipologías —significado «tipológico» o categorial— (Lemke, 1998), mientras que los matices referidos a grados se representan más fácilmente mediante gestos motores o figuras visuales, conocidos como «significado de variación continua» o «topológico» (Manghi, 2009). Por ejemplo, en la lengua natural no poseemos una forma estándar para expresar de forma precisa grados intermedios de las categorías existentes; esto lo realizamos mediante gestos o dibujos (Lemke, 2002).

Halliday (1978) consideró los signos lingüísticos funcionalmente como recursos para producir significados. Cada texto podía significar según *presente* (idea), *oriente* (relación) y/u *organice* (estructura, complejidad textual, precisión...) los datos. Más tarde, desde la Teoría Semántica Social, Kress y Van Leeuwen (2001) completaron esta propuesta con una gramática que atribuía significados a cada modo según tres metafunciones o prácticas:

1. **Discursiva (presentación):** Representar y comunicar *ideas, hechos* (el tema).
2. **Social (orientación):** Representar la *relación interpersonal* entre las actitudes del emisor y del receptor respecto al mensaje (contexto, emociones, transitividad, posición, perspectiva...).
3. **Textual (organización):** Hacer posible la producción de mensajes con una estructura lingüística *textual* (léxico, gramática, fonología, tipografía, grafología, escenas, episodios, figura, partes, color, encuadre, sonido, música...).

> Tipos de estudios multimodales

En el análisis de los sistemas multimodales se suelen aplicar diversos enfoques, los cuales están relacionados con el tipo de fusión de los datos multimodales de entrada (Wahlster, 2002). En general, se pueden considerar tres enfoques para realizar la evaluación a nivel de sistema: el *experimental*, el *predictivo* y el *basado en expertos*. Para el análisis que nos interesa, distinguiremos dos vertientes en los estudios multimodales: la intrasemiótica y la intersemiótica (Manghi, 2009).

> Intrasemiosis

Entre estas investigaciones encontramos las de Van Leeuwen para el cine; Kress y Van Leeuwen para el diseño visual; Halliday para la lengua; O'Toole para las artes visuales; Steiner y Van Leeuwen en música y sonido; O'Toole, Bardry y Thibault, Galán, D'Angelo et al., Stenglin y Ravelli en arquitectura y artes visuales; Martinec en acción y gestos; Van Leeuwen para el significado tipográfico; y O'Halloran para el simbolismo matemático.

Modos de audición en los audiovisuales

De modo general, Kress y Van Leeuwen (2001) plantean múltiples articulaciones en todos y cada uno de los signos, en cualquiera de los sistemas semióticos. Desde su propuesta, el estrato del *contenido* podría analizarse en *discurso* y *diseño*, mientras que el estrato de la *expresión* se podría analizar en *producción* y *distribución*. A continuación podemos ver este ejemplo aplicado al lenguaje y a las imágenes visuales:

LENGUAJE	CONTENIDO	DISCURSO	Idea
		DISEÑO	Aspecto conceptual de la expresión, gramática, forma, frases, melodía, ritmo...
	EXPRESIÓN	PRODUCCIÓN	Escritura, sonido, timbre, interpretación, ejecución.
		DISTRIBUCIÓN	CD, vídeo, en directo.

IMAGEN	CONTENIDO	DISCURSO	Relaciones intervisuales, trabajo.
		GRAMÁTICA	Escena, episodio, figura, parte...
	EXPRESIÓN	GRÁFICOS	Sistemas multifuncionales

A modo de ejemplo, y en el caso de la escucha sonora-musical, otro de los planteamientos posibles es el del laboratorio Cresson (Porta y Ferrández, 2009), que propone tres ámbitos de la observación de la escucha: la *descripción* (qué se escucha), el *recorrido* (en qué momento se escucha), y la *experiencia* (desde dónde se escucha).

> Intersemiosis

La segunda vertiente del análisis multimodal reconoce que la complejidad semiótica tiene que ver con las relaciones entre los diversos recursos semióticos utilizados para crear significado (Manghi, 2009). Cuando dos o más modos son coarticulados, tienen el potencial de generar una sinergia mayor que

la suma de sus significados aislados (Lemke, 1998). Por ejemplo, los gestos y las palabras son dos modalidades no redundantes, sino coexpresivas (Quek et al., 2002).

En el terreno de los estudios intersemióticos debemos citar a Baldry y Thibault en productos audiovisuales, Lemke en hipertexto o Lazarus en psicoterapia cognitivo-conductual basada en información global del paciente (biológica, afectiva, sensitiva, imaginativa, conductual, cognitiva y social). A partir de toda esta información, se plantean enfoques en niveles supuestamente cada vez más eficaces, con la estrategia de analizar la conducta clave, lo que concuerda con el análisis de los llamados *memes* o unidades de información significativas a nivel estilístico o cognitivo, que permiten reproducir conceptos o culturas (Blackmore, 2000).

En el enfoque intersemiótico destacaremos tres conceptos que nos serán útiles para nuestro trabajo:

1. **La complementariedad intersemiótica (*intersemiotic complementarity*):** Fue inicialmente propuesta por Royce (1998) para analizar cómo los modos verbal y visual se complementan semánticamente para producir un único fenómeno textual. En este sentido se incluye la repetición, la sinonimia, la antonimia, la metonimia, la hiponimia, etc. Cabe distinguir la posible convergencia (cohesión) o divergencia de significados entre modos en el mismo mensaje (O'Halloran, 2008).
2. **Las transiciones intersemióticas:** Se refieren a los procesos de traducción o traslación de contenidos entre diversos modos semióticos, entre los que destaca la metáfora.
3. **La resemiotización:** Cambio de significado de contexto a contexto o de práctica social a práctica social (Iedema, 2003). Es muy adecuada para estudios interculturales o de patrimonio histórico.

> Intersemiosis audiovisual: la imagen y el sonido

El audiovisual es fundamentalmente audio y vídeo. Uno de los interrogantes más antiguos en psicología experimental concierne a la naturaleza de las interacciones entre modalidades sensoriales, es decir, las posibles prioridades o jerarquías entre los sentidos cuando proveen información contradictoria. La in-

Modos de audición en los audiovisuales

formación visual y auditiva pueden reforzarse mutuamente, contradecirse o modificarse una a la otra (Vines et al., 2006).

Sabemos que la visión influencia la audición: la música a lo largo de la historia (antes de la radio y el disco) siempre había sido escuchada viendo la ejecución de los intérpretes, que mediante sus gestos y expresiones faciales señalaban la concentración, la emoción, la segmentación, las dificultades, las disonancias, los saltos, las expectativas, etc. (Thompson, Graham y Russo, 2005).

La audición influencia también la visión, debido sobre todo a una cierta asimetría cerebral en la representación del espacio. A pesar de que los eventos auditivos se perciben más rápidamente que los visuales (Spence y Driver, 1997), la gente tiende a advertirlos como eventos simultáneos, y como el audio precede el vídeo, este queda influenciado por el primero (Vines et al., 2006). Al querer explicar este fenómeno, muy habitual en los audiovisuales, Neumann, Van der Heijden y Allport (1986) señalaron posibles diferencias entre la atención auditiva y la visual: los eventos auditivos en el mundo real tienden a ser intermitentes y transitorios, en comparación con los eventos visuales, más continuos. Por lo tanto, tendría sentido evolutivo organizar cualquier dependencia multimodal de manera que la atención visual se desplace hacia eventos auditivos, y no a la inversa, ya que los eventos de interés son más propensos a volverse inaudibles, aunque sigan siendo visibles, que a convertirse en invisibles pero seguir siendo audibles.

Otra posible explicación proviene del ámbito de la emoción percibida, evitando una correspondencia perfecta entre la emoción provocada visualmente y la provocada auditivamente, a pesar de un cierto grado de redundancia entre ambos modos de comunicación (Vines et al., 2006).

> La música y el movimiento en los audiovisuales

Algunas analogías entre parámetros musicales y del movimiento visual se han propuesto para el análisis audiovisual: el tempo con la velocidad, los cambios de altura con ascender o descender en el espacio, los cambios de intensidad con cambios en la distancia o con los niveles de energía que activan el movimiento, etc. Los cambios de tono están fuertemente asociados a los descensos espaciales —a modo de efecto Doppler (Neuhoff y McBeath, 1996)—, mientras que los aumentos de intensidad están más asociados con el aumento de velocidad que con movimiento vertical (Eitan y Granot, 2006). En este sentido, Collier

y Hubbard (2001) encontraron cierta interacción entre altura y tempo: las aceleraciones fueron percibidas como más rápidas en tonos agudos y las desaceleraciones como más notables en tonos graves.

La relación entre altura y verticalidad no es sencilla. La notación occidental señala frecuentemente que su principal factor es la verticalidad respecto a la altura; hay raíces culturales e históricas, pero no biológicas para ello. Cox (1999) explica la verticalidad respecto a la altura del sonido debido a que las notas agudas requieren grandes cantidades o magnitudes de aire, esfuerzo y tensión. La localización de altura/verticalidad es aprendida más que innata: los preescolares pocas veces usan los términos «alto» o «bajo» para referirse a agudo y grave (Hair, 1981). En cambio, sorprendentemente, la altura auditiva ha sido asociada a la posición lateral, de modo que los sonidos más agudos se encontrarían a la derecha, y los graves, a la izquierda, como en el piano (Mudd, 1963).

Los movimientos corporales parecen revelar de forma precisa estados mentales internos y actitudes, sugiriéndose que los gestos y movimientos de los músicos comunican aspectos de su estado interno y de su intención expresiva, tanto más que los sonidos (Davidson, 1993). La relación entre curvas de intensidad y gestos dinámicos se refleja en las «formas esenciales» de Clynes y Nettheim (1982), iconos arquetípicos de las emociones humanas básicas, que incluyen tono, volumen, movimiento y curvatura visual. Clynes demostró que las emociones asociadas con formas esenciales eran identificadas correctamente en distintas culturas cuando estas formas se presentaban a través del sonido, la visión o la presión táctil. Los resultados de Clynes revelan la capacidad de asociar contornos de intensidad de forma intermodal.

De hecho, en estudios donde los oyentes señalan el grado de tensión de una música, Krumhansl y Schenck (1997) demostraron que tanto los que escuchaban una pieza musical como los espectadores de una coreografía compuesta sobre esa música estuvieron de acuerdo en sus atribuciones de tensión, incluso cuando los oyentes no hubiesen visto el baile ni los espectadores hubiesen escuchado la música. Así pues, las fluctuaciones de tensión transmitidas por los parámetros del movimiento se correlacionan con los parámetros musicales. Los espectadores solo visuales o solo auditivos identificaron la mayoría de finales de frases. Tanto la información auditiva como la visual coinciden no solo en la tensión, sino también en la estructura.

A modo de resumen, el siguiente cuadro refleja las principales relaciones entre aspectos del sonido y del movimiento (adaptado de Eitan y Granot, 2006).

Modos de audición en los audiovisuales

		Dirección vertical		Cambio de distancia		Dirección lateral		Cambio de velocidad		Fuerza externa	
		Ascendente	Descendente	Acerca	Aleja	Derecha	Izquierda	Más rápido	Más lento	Aporte	Niega
Dinámica	Crescendo			x				x		x	
	Diminuendo		x		x						
Contorno melódico	Ascendente							x			
	Descendente		x				x				
Intervalo inicial	Acelerando							x			
	Decelerando								x		
Paso	Acelerando							x			
	Decelerando		x								
Articulación	Tenuto										
	Staccato								x		

> Multimodalidad y educación audiovisual

Moreno y Mayer (2000) presentan los siguientes principios de la teoría cognitiva, procedentes de la multimodalidad de los audiovisuales, que deberían ser considerados y aplicados en los aprendizajes multimedia:

1. **Principio multimedia:** Se aprende con mayor eficacia de palabras asociadas a imágenes que de palabras solamente.
2. **Principio de contigüidad espacial:** Se aprende más eficazmente cuando las palabras y las imágenes se corresponden y se presentan una junto a la otra en la pantalla, en lugar de estar alejadas.

3. **Principio de contigüidad temporal:** Se aprende más eficazmente cuando las palabras y sus imágenes correspondientes son presentadas al mismo tiempo, en lugar de por separado.
4. **Principio de modalidad:** Se aprende más eficazmente cuando el material multimedia combina animación y narración sonora que cuando ofrece animación y texto en pantalla.
5. **Principio de redundancia:** Se aprende con mayor eficacia de la animación y la narración sonora combinadas que de la animación, la narración y el texto en pantalla.
6. **Principio de las diferencias individuales:** Los estudiantes con conocimientos bajos prefieren efectos de diseño más impactantes que los que tienen un conocimiento alto.

Por otra parte, el progreso explosivo de la tecnología informática ha fomentado la tendencia a presentar materiales didácticos mediante proyectores y portátiles. Las conferencias combinadas con multimedia pueden ser mejores para algunos estudiantes, pues les permiten construir conocimientos significativos y hacer conexiones referenciales de representaciones mentales (Mayer, 2001; Paivio, 1986).

Por lo tanto, una alfabetización tecnológica multimodal debe proporcionar el conocimiento de los lenguajes, así como los medios necesarios para el análisis y la producción de mensajes a través de herramientas digitales (cámaras, ordenadores, etc.) (Bautista, 2007). La música y el sonido son modos eficaces de comunicación, con un alcance emocional y motivacional que no debe pasarse por alto. No lo olvidemos.

> Epílogo

A modo de síntesis, presentamos a continuación un ejemplo de tabla que permitirá un análisis multimodal de los audiovisuales: