

2.8 El sonido en el cine

Como ampliamente describe Noël Burch (1995) en su obra *El tragaluz del infinito*, cuyo objetivo confeso es demostrar que “el lenguaje del cine no tiene nada de natural ni de eterno, que tiene una historia y que está producido por la Historia”, el nacimiento del cine fue recibido como la respuesta a lo que siempre había sido una de las más preciadas aspiraciones burguesas: la posibilidad de recrear la vida.

Refiriéndose al interés de Edison de grabar y reproducir imágenes asociadas a su fonógrafo para disfrutar de óperas cuyos protagonistas ya habían muerto, dice Burch (1995:22):

“..no solamente es la ambición de un hábil líder industrial, sino también la búsqueda del fantasma de una clase convertido en el de una cultura: llevar a cabo la <conquista de la naturaleza> triunfando sobre la muerte de un *ersatz** de la Vida Misma”. (*sucedáneo)

Este afán “frankensteiniano” respecto del naciente cine, del que Burch excluye como partícipes intencionados a los Lumière, más técnicos y científicos que artistas o comerciantes, se refleja simplemente en los nada casuales nombres que recibieron los incipientes aparatos cinematográficos, tales como bioscopio o vitascopio. Bajo esta “neurosis naturalista” se desarrolló el aparato cinematográfico de modo que cada uno de sus avances tecnológicos a lo largo de la historia han ido siendo presentados como la restitución de una carencia.

El reto de plasmar en el invento lo que Burch llama “la ideología naturalista de la representación, dominante en el momento en casi toda la *intelligentsia* europea” (Burch 1995:41) encontró su primer obstáculo en la posibilidad de dotar al cine, no ya de sonido sincrónico, sino, directamente y en primera instancia, de “alma”, es decir, de palabra.

Diversos aparatos tales como el fonoscopio (asociación de fonógrafo y fonoscopio en proyección, obligados a funcionar de forma sincrónica) que Demeny propuso a Gaumont en 1895 y más tarde el cronófono-Gaumont (1900) permitieron la creación del primer cine

hablado antes de 1910. Curiosamente, para Edison la posibilidad de reproducir “fotografías en movimiento” venía en cambio a llenar una carencia (la reproducción de imágenes) de su fonógrafo y hacia ello dirige sus investigaciones posteriores que darán como fruto el Kinetoscopio Parlour. (Burch 1995).

Como quiera que diversos condicionantes técnicos y, principalmente, comerciales frenaron el pleno desarrollo del cine sonoro hasta a partir de 1928, lo que nos interesa resaltar es la vocación primigenia del cinematógrafo de constituirse en un reproductor “naturalista” de la realidad, merced a la ideología que Burch denuncia como “la visión institucional”.

En este sentido e independientemente del conocido debate de la época entre partidarios y detractores del cine sonoro, tres ideas resultan fundamentales para nuestro propósito:

- El cine nunca fue silente.
- Antes que dotar al cine de sonido en sentido estricto, se pensó en dotarlo de palabra.
- Desde que el cine es sonoro, el cine es vococentrista.

La primera aseveración se basa en la reflexión de que incluso las primeras películas fueron pensadas para ser proyectadas acompañadas de música, bien previamente grabada o bien interpretada en directo, aunque sólo fuese para ocultar el ruido que producía el proyector. La figura del explicador o comentarista, fue otro recurso para solventar su carencia de sonido, principalmente de voz.

En este sentido, Camper (1985:369) acuña el término de películas “silentes con sonido”, Jean Painlevé escribió: “el cine siempre ha sido sonoro”; Bresson precisa “nunca ha existido el cine mudo” (citados ambos por Chion 1982).

La segunda afirmación se deriva de que la primera preocupación para dotar al cine de su deseado realismo o naturalismo, consistió, sin duda, en dotarlo de voz. De ahí los afanes de Edison y compañía; la temprana actuación de los actores y actrices del primer cine, que simulaban hablar, aún cuando supieran a ciencia cierta que no serían escuchados; o los rótulos e intertítulos que también vinieron a dar respuesta a esta necesidad. Antes de las posibilidades

comerciales y de la expansión del cine sonoro como tal, gracias al cronófono de Gaumont, Alice Guy rodaría entre 1900 y 1907 más de 100 cintas cantadas o habladas (Burch 1995:44). La denominación de “*talkies*” para las primeras películas sonoras, evidencian también esta afirmación.

Es en este sentido, Chion (1982) afirma que si consideramos la gestualidad de sus actores, el constante movimiento de sus labios, más que ser mudo, el cine era sordo.

“La duda entre mudo y silencioso, como aquella entre hablado y sonoro, gira en torno a un mismo centro, la palabra, la voz.”(Chion 1982:16)

En un sentido semejante Altman (1980) demuestra que a pesar de la insistencia por parte de los teóricos para crear una relación de contrapunto entre sonido e imagen, el cine clásico narrativo sigue siendo un cine de palabra sincronizada, donde, salvo algunas excepciones, los actores ganan su derecho a aparecer en pantalla en la medida en que hablan. “Hablo, luego soy visto” (Altman 1980:68). Es por ello que el *découpage* clásico es el de plano - contraplano en el cual la cámara enfoca alternativamente a cada hablante en una situación de diálogo. No en vano los franceses llaman “*cinéma parlant*” al cine sonoro.

Para Altman, “la edición de sonido es primero y principalmente, edición de palabra; su propósito primordial es asegurar la claridad de los diálogos cinematográficos” (Altman 1980:68)

Volveremos sobre el papel de la voz en el cine, sirvan de momento estas tres ideas como preámbulo a lo que el sonido ha significado para el cine.

2.8.1 Características y dimensiones del sonido cinematográfico

El sonido en el cine puede presentarse bajo tres formas: música, ruidos (o efectos sonoros) y voz. En contraste, más adelante consideraremos el silencio, cuya existencia expresiva solamente es posible en el cine cuando éste es sonoro.

Las tres propiedades acústicas de cualquier sonido, a saber: intensidad (referido a la amplitud de las vibraciones que el sonido produce en el aire); tono (frecuencia de estas vibraciones, medidas en hertzios) y timbre (característica o color del sonido que ayuda a la identificación de su fuente), son manipuladas por el cine para conseguir efectos precisos, bien sean estéticos, dramáticos o de inteligibilidad. Esta manipulación es posible merced a la dislocación o asincronía en la producción del sonido respecto de la imagen y al fenómeno de la acusmatización, referido a la voz, del que nos ocuparemos más adelante.

De esta manera el sonido cinematográfico permite y se caracteriza por las cuatro dimensiones que Bordwell y Thompson (1985; 1992;1995) identifican: ritmo, fidelidad, espacio y tiempo, que a continuación comentamos.

1. El sonido tiene una duración y, por lo tanto un **ritmo**, un tempo, un compás. Lo más frecuente es adaptar el ritmo de la música y de los ruidos a la imagen. Creemos que dentro de esta dimensión rítmica se incluyen conceptos como el de contrapunto, principalmente referido a la música que expresa ideas o sentimientos distintos a los de la imagen, pero también puede referirse a los ruidos cuya fuente visible es otra distinta a la esperada.
2. La **fidelidad** del sonido cinematográfico está referida no a la coherencia con la fuente que lo produce, sino con la que la imagen y la narración proponen como fuente.
3. El sonido tiene una dimensión **espacial**. Esta dimensión está referida principalmente a la capacidad del sonido cinematográfico para reconstruir un

paisaje. No solamente mediante la manipulación de la intensidad del sonido en combinación con la distinta escala de planos, o profundidad de campo de las imágenes, sino más allá, en la construcción del espacio diegético, gracias a la voz en *off*. También gracias a los sonidos *over* de los que se valen los sonidos internos o subjetivos, como los pensamientos de los personajes, el sonido permite el acceso a espacios, como la mente, vedados para la imagen. De manera semejante a como las imágenes bidimensionales de la pantalla producen para el espectador un efecto tridimensional, el rectángulo de la pantalla se amplifica y cobra relieve, gracias al sonido cinematográfico.

4. Por último, el sonido tiene una dimensión **temporal**. La banda sonora puede producir sonido que expresa un tiempo distinto al de la imagen, anterior o posterior. Dentro de esta dimensión podemos ubicar conceptos como los de sonido sincrónico (el que se produce al mismo tiempo que el que la imagen sugiere) o asincrónico (entendido como la diferencia en el tiempo entre lo que se ve y lo que se oye).

La identificación de los distintos casos de cómo el sonido se produce en la práctica cinematográfica, ha sido objeto de numerosos estudios y análisis diversos. Sin embargo, el concepto de diégesis como historia que cuenta la película, ha permitido la clasificación de distintos usos del sonido en el cine. Sin pretender que sean exhaustivas, ni únicas en cuanto a los conceptos que emplean, las siguientes tablas propuestas por Bordwell y Thompson (1985 y 1995) nos permiten identificar los casos más frecuentes:

Sonido cinematográfico respecto a su dimensión espacial

Diegético (dentro del espacio de la historia)		No diegético (fuera del espacio de la historia)	
En pantalla		En off	
Interno (subjetivo: no lo oyen personajes) También llamado "over"	Externo (objetivo: lo oyen los personajes)	Interno (subjetivo: no lo oyen los personajes) También llamado "over"	Externo (objetivo: lo oyen los personajes)
Procede de una fuente externa a la historia. Igual que sucede con el narrador omnisciente. Por decirlo de alguna manera, siempre es "interno, subjetivo". No lo oyen los personajes.			

Tabla 1

Relaciones temporales del sonido cinematográfico respecto a la diégesis

Relación temporal	Espacio de la fuente	
	Diegético (Espacio de la H ^a)	No diegético (Espacio fuera de la H ^a)
1. No simultáneo: sonido <i>anterior</i> en la historia que en la imagen.	<i>Flashback</i> sonoro; <i>flashforward</i> en la imagen; solapado del sonido. (Ej. Vemos verja rota y oímos sonido de accidente de coche, sin pistas de que alguien esté recordando)	Sonido señalado como pasado sobre la imagen. (Ej. Sonido de un discurso de Kennedy sobre imágenes de USA de hoy)
2. Sonido <i>simultáneo</i> en la historia y en la imagen.	<i>Externo</i> : diálogos, efectos, música. <i>Interno</i> : pensamientos del personaje que se oyen.	Sonido señalado como simultáneo con las imágenes sobre la imagen. (Ej. Un narrador describiendo los hechos en tiempo presente).
3. No simultáneo: sonido <i>posterior</i> en la historia que la imagen.	<i>Flashforward sonoro</i> ; <i>flashback de imagen</i> con sonido que continúa en el presente; el personaje narra hechos anteriores; solapado de sonido. (Ej. El testigo en el juicio cuenta mientras hay <i>flashback</i> en imágenes de lo ocurrido. Un narrador que recuerda)	Sonido señalado como posterior sobre la imagen. (Ej. Un narrador que recuerda lo que las imágenes muestran).

Tabla 2

De lo hasta aquí expuesto nos interesa destacar dos reflexiones.

La primera: el cine y especialmente el cine narrativo de ficción, está formado por materiales heterogéneos, todos ellos manipulables y de hecho, manipulados, para conseguir unos determinados efectos de naturalismo y realidad en sus espectadores.

La segunda: las dos bandas, la sonora y la de imágenes, perfectamente separadas en el momento de su producción, manipulables cada una de ellas conforme la intención del discurso que la película genera, consiguen en su reunificación, que se materializa en el momento de su proyección-emisión, los deseados efectos de realidad. En este sentido, resulta interesante comprobar que incluso la teorización de las dimensiones del sonido cinematográfico arriba señaladas y otras semejantes, parten de la premisa de este fenómeno.

Es decir, el sonido cinematográfico no puede analizarse ni definirse sino bajo la realidad de su acusmatización y posterior relación con la banda de imágenes. Y es bajo esta relación con las imágenes que el sonido cinematográfico cobra un sentido específico.

2.8.2 La voz en el cine

Chion (1982), uno de los teóricos del cine que más ha estudiado la dimensión sonora del lenguaje cinematográfico, afirma que el cine es principalmente vococentrista. Define esta categoría mediante el análisis del papel de la voz en el cine en ambos sentidos: desde la emisión y desde la recepción.

Por una parte, como arriba apuntábamos, desde su nacimiento el cine sonoro pone a punto todas sus sustancias sonoras con el fin último de destacar la voz humana, principalmente su inteligibilidad (“este vococentrismo es, casi siempre, verbocentrismo”). Lo que se llama “toma de sonido” es en realidad la captación de las voces. Normalmente el resto de sonidos que se producen durante el rodaje son eliminados o reelaborados subsidiariamente a su función e intención narrativa y en beneficio de las voces.

Por otra parte, la reflexión de Chion en torno al vococentrismo del cine, incluye la característica humana de percibir, de entre todos los sonidos de su entorno, preferentemente la voz humana.

“...del mismo modo que la madre se despierta cuando los lloros lejanos de su hijo alteran el ruido habitual -a menudo más intenso- de la noche, está, en el torrente de sonidos, en primer lugar ese otro «nosotros» que es la voz de otro que se dirige a nuestra atención.” (Chion 1982:16).

Y más tarde:

“Llamamos “vococentrismo” al proceso por el cual, en un conjunto sonoro, la voz atrae y centra nuestra atención, del mismo modo que el rostro humano, en un plano cinematográfico, atrae la atención de nuestra vista” (Chion 1998:282-183)

De igual forma que cuando frente a una figura que recuerde la humana lo primero que percibimos visualmente es la cara, en el terreno de la percepción auditiva, es la voz lo primero que las personas percibimos de nuestro entorno sonoro.

“Parafraseando a Christiane Sacco, en *Plaidoyer au Roi de Prusse*: «*La presencia de un cuerpo estructura el espacio que le contiene*» (se refiere, por supuesto, al cuerpo humano) ‘*la presencia de una voz humana estructura el espacio sonoro que la contiene.*’ (Cursivas del autor: Chion 1982: 15)

En una obra posterior Chion (1998:98) recuerda que no es casual que “la palabra griega *phoné*, de la que se han sacado todos los nombres relativos a la grabación sonora, quiera decir voz”.

Además de reconocer la voz de entre todos los sonidos, la percepción humana actúa en el sentido de intentar conferirle un sentido cuando esta voz habla en una lengua que le es conocida “sin pasar a la interpretación de los demás elementos hasta que esté saturado su interés por el sentido”. (Chion 1990 b: 18)

En esta dirección, explicando las motivaciones de las que se deriva el placer de escuchar, Doane (1980) señala que durante la recepción del diálogo sincrónico y de la voz en *off* en las películas de ficción, el espectador experimenta el placer de escuchar sin ser visto ni oído, placer que apela al deseo denominado por Lacan (1977) “instinto invocatorio”. “Esta actividad, con respecto a la banda sonora, no es distinta al voyeurismo frecuentemente explotado por la imagen cinematográfica” (Doane 1980:169)

Más allá del efecto realista que el sonido sincrónico produce, más allá de la inteligibilidad y del placer de percibir una voz con su volumen, ritmo, tono y timbre, Doane apela al psicoanálisis para recordar que, en la medida en que el espacio del niño está definido en términos de lo audible y no de lo visible, en la medida que la voz tiene un mayor mandato

sobre el espacio que la vista, “para el niño, la voz, incluso antes del lenguaje, es el instrumento de la demanda” (Doane 1980:169).

La voz de la madre, es el mayor componente sonoro que envuelve al niño en la primera infancia. Además, a diferencia de la vista, la voz permite un juego reversible gracias al cual el sonido es simultáneamente emitido y oído por el sujeto. Opuesto a la situación de ver, es como un “espejo acústico” en permanente funcionamiento. Si además consideramos que podemos oír sonidos internos al cuerpo o a través de objetos opacos o detrás de nosotros, encontramos dos juegos de términos puestos en oposición: exterior/enfrente/vista e interior/espalda/escucha, (Rosalto 1974: citado por Doane).

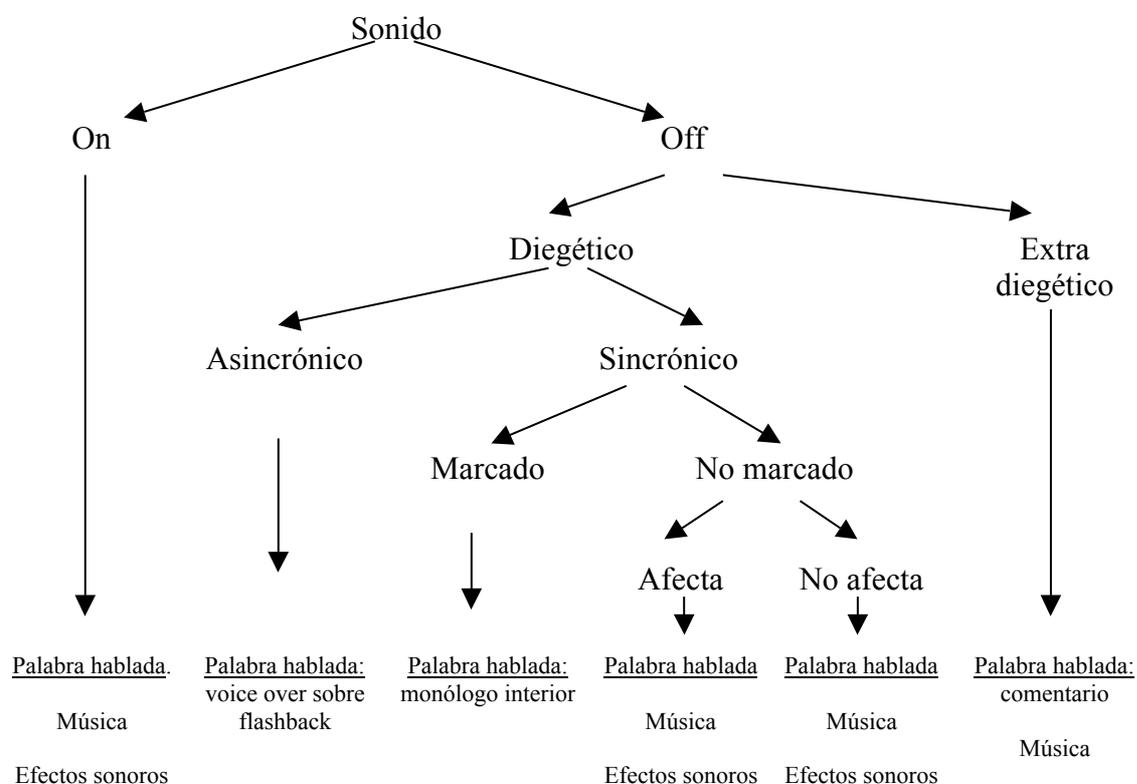
Por otra parte, Daniel Percheron (1980:16-23) realiza un análisis exhaustivo del sonido cinematográfico con el fin de conseguir una nueva taxonomía. Para ello identifica diversos puntos de vista a partir de los cuales el sonido puede ser considerado y que a continuación explicamos:

1. Dependiendo de su relación con la banda de imágenes: el sonido puede ser *on*, si su causa aparece identificada en imagen u *off* si no aparece la fuente en imagen. Además, acuña el término “*off-on*” para casos como el monólogo interior en que el personaje aparece en pantalla pero no pronuncia las palabras que se escuchan y que proceden de él mismo.
2. Dependiendo de su relación con la diégesis: el sonido puede ser diegético si su fuente es también fuente identificada dentro de la diégesis o puede ser extradiegético (entendiendo “extra” como algo añadido más que “desde fuera”). Dentro de esta última categoría Percheron distingue únicamente la música y el comentario voice-over. Pero distingue dos tipos distintos de diégesis: llama “no marcados” (*unmarked*) a los filmes cuya diégesis oculta la actividad narrativa; y llama “marcados” (*marked*) aquellos que colocan en primer plano el acto de narrar. Por otra parte, identifica que en los primeros, el sonido *off* puede o no afectar a lo que se dice o a lo que se ve en la imagen.

3. La forma como el espectador recibe el sonido. Así, distingue sonidos de transmisión diegética (el sonido pasa del film al espectador filtrado a través de un personaje, es decir, lo que oye el espectador también lo debe oír el personaje) y sonidos de transmisión extradiegética (el sonido pasa directamente desde el film al espectador sin ser mediado por un personaje).

4. Finalmente distingue entre sonido sincrónico y asincrónico, dependiendo de la relación temporal del sonido con la banda de imágenes pero, especialmente, con la diégesis.

La conclusión más importante de este análisis consiste en identificar, de entre todas las sustancias sonoras, la palabra hablada, la voz, como “el material significativo dominante y principal componente del sonido en el cine”. Esta conclusión se aprecia con evidencia en el siguiente diagrama que ilustra el análisis de Percheron (1980:23):



El autor resume así su trabajo: “Los efectos sonoros y la música son raramente portadores de significado. Los efectos sonoros, la mayoría de veces, son sólo articulaciones añadidas de la banda de imagen que ayudan a expandirla y a hacerla diegética. Es bien conocida la más bien limitada y vulgar función enfática y la programática de la música cinematográfica, rol éste último que se le asigna a dicha música nueve de cada diez veces. Escapando de estos semánticamente empobrecidos tipos de sonido, confrontando y trabajando con la banda de imagen, solamente la palabra hablada puede constituir un sistema altamente significativo”. (Percheron 1980:23)

2.8.2.1 Los distintos tipos de voz cinematográfica y su relación con la diégesis y el espectador.

Hemos visto que al referirse a la dimensión espacial del sonido, autores como Bordwell y Thompson, entre otros, recurren al concepto de diégesis, como referente indispensable para el análisis o la taxonomía. Pero la diégesis solamente cobra sentido en relación con el receptor. Es el espectador de cine el que construye, el que hace realidad, el concepto de diégesis. Es en esta imbricación de distintos componentes del fenómeno, cuando la diferencia entre conceptos como emisor y receptor parece diluirse, que el análisis se complica.

Así, la voz en el cine puede ser: voz sincrónica o *in*, voz en *off*, voz *over*.

Por **voz sincrónica** (*in*, u *on*) entendemos aquella cuyo presunto emisor aparece en la pantalla en el momento en que habla. Algunos autores la denominan, por eso, voz vinculada (Chateau/Jost 1979). A pesar de los numerosos experimentos de otro tipo de relación (Clair, Lang, Vigo, Godard, Straub, Duras, Kubelka etc.), la voz sincrónica sigue siendo la forma dominante de representación sonora en el cine. Su dominancia estriba en la necesidad de proveer al cuerpo que la contiene (llamado fantasmático *-fantasmatic-* por Doane:1980, o considerado un mero pelele de ventrilocuó por Altman:1980) de una apariencia de completud de la cual carece.

Pero sobre todo y fundamentalmente, como el resto de los sonidos y de los avances técnicos en su grabación, la voz sincrónica está llamada a “disminuir el ruido del sistema, ocultando el trabajo del aparato, y así reduciendo la distancia percibida entre el objeto y su representación” (Doane 1980:164) A diferencia de la voz en *off* que puede suponer un riesgo para la apariencia de homogeneidad de la película, la voz sincrónica ata de manera perfecta todos los componentes.

Por **voz en *off*** entendemos aquella que se produce cuando el presunto emisor no está en pantalla. El uso más frecuente de la voz en *off* consiste en que el presunto emisor está contenido en la diégesis pero fuera del cuadro. Se trata de un hablante que, antes o después, será identificado por el espectador. Desde el punto de vista de la teoría filmica, la voz en *off* responde a la necesidad de recrear una dimensión espacial lateral. Permite así, la reconstrucción del espacio diegético que la imagen, confinada en los límites del cuadro, no puede por sí sola conseguir. “El uso tradicional de la voz en *off* constituye una negación del cuadro como límite y una afirmación de la unidad y homogeneidad del espacio representado” (Doane 1980:165)

El empleo mismo del término ha sido cuestionado por distintos autores. Bailblé (citado por Doane 1980:165) argumenta que la voz en *off* debe de ser siempre una voz *in* porque la fuente literal del sonido en la sala es siempre un altavoz emplazado detrás de la pantalla (o, en todo caso, añadimos, dentro de la sala y por tanto, “*in*” para el espectador). El término indica también una teorización que parte de una presunta predominancia de lo visual. Predominancia no específica del cine, sino, de acuerdo con Metz (citado por Doane 1980), de una cultura. Por eso Metz argumenta que el sonido nunca está *off* porque siempre es audible.

Por **voz *over*** entenderemos aquella voz cuyo emisor puede o no ser conocido por el espectador, dependiendo del género y de la intención de la diégesis, pero que, en todo caso, su imagen permanece fuera de pantalla, excepto en el monólogo interior, en que la imagen del “emisor” es mostrada sin articular las palabras. Es quizás por ello que frecuente algunos teóricos llamen indistintamente voz *over* a la voz en *off* y viceversa. Kozloff (1988:5) la define como “enunciados orales que vehiculan cualquier porción del relato, pronunciados

por un locutor invisible, situado en un espacio y un tiempo que no sean los que se presentan simultáneamente a las imágenes que vemos en la pantalla”.

Por último, el único uso del concepto voz *under* es aquel referido a la traducción simultánea. Se caracteriza porque solamente las primeras líneas intentan estar en sincronía con el movimiento de los labios y se produce manteniendo en segundo plano el sonido de la voz que habla en el idioma original.

Una de las aportaciones más interesantes sobre el sonido cinematográfico de entre las que hemos tenido acceso, es la identificación de los tres espacios diferenciados del sonido cinematográfico que Doane (1980:166) distingue y su relación con los distintos usos de la voz en el cine. A saber:

- 1) El espacio de la diégesis.
- 2) El espacio visible de la pantalla como receptora de la imagen.
- 3) El espacio acústico del teatro o auditorio.

Mientras que el primero no tiene límites físicos, no puede ser medible y constituye el espacio virtual construido por el film, contenido en sustancias audibles y visibles (así como en la implícita y también virtual posibilidad de que sus objetos pueden ser tocados, oídos o gustados), **el espacio visible** está contenido en la **pantalla** medible (porque estrictamente hablando, la pantalla no es audible, lo son sus altavoces) y, en cambio, **el espacio acústico** del teatro, cuyos límites son más extensos que los de la pantalla, envuelve al espectador.

Paradójicamente, nada une a los tres espacios, salvo la práctica significativa de la película en sí misma y el hecho físico de encontrarse todos en la misma sala. Es en el primer espacio, en el de la diégesis, en donde los personajes existen y a partir del cual los espectadores pueden dar fe de su existencia. Pero los tres espacios se relacionan de manera constante y compleja, algo que se hace patente, por ejemplo, en la voz en *off*.

La relación de los distintos tipos de voz con estos espacios depende del género cinematográfico (Doane 1980). El cine clásico narrativo trabaja para negar los dos últimos espacios. Por eso la voz sincrónica es dominante y preferida para este efecto; de ahí el tabú en el género que impide que un personaje mire a la cámara y “hable” al espectador. La voz en *off* trabaja también en función de un cuerpo, el de un personaje que está confinado en algún lugar de la diégesis y que podrá ser visto con un mero reencuadre que reunificará la voz con su cuerpo.

En la medida en que tanto la voz en *off* como la voz en *over* hacen patente, aunque sea por unos segundos (como sucede en el solapado de sonido de/o en los planos-contraplanos de algunos diálogos) el fenómeno de la acusmatización, es decir, la invisibilidad de la fuente, ambas son inquietantes y misteriosas, de ahí su profusión en el cine negro y de terror. Dicen Gaudreault y Jost (1990:82) que “cuando un filme empieza con una voz *over* nuestra primera inquietud consiste en desacusmatizarla (...) es decir, en encontrarle una fuente visual *in* o vinculada, dado que esta operación nos permitirá determinar el lugar del narrador explícito en relación a la diégesis”. Reflexionando cómo el sonido es el camino ideal para explicar lo misterioso, lo invisible, lo sobrenatural, igualmente Altman (1985:73) concluye: “el sonido siempre acarrea consigo la tensión de lo desconocido hasta que es anclado por la vista”.

Para Doane (1980) en el cine narrativo el cuerpo actúa siempre como un soporte para la voz, incluso para la voz en *over* durante un *flashback* o el monólogo interior. Aunque durante el *flashback* aparezca una dislocación temporal respecto al cuerpo, la voz es frecuentemente retornada al cuerpo como forma de cierre. En el monólogo interior la voz *over* permite el acceso al interior del cuerpo que se presenta como su continente y, de esta manera, excede lo visible.

Algo distinto ocurre con la voz *over* en los géneros documentales, porque se trata de una voz descorporizada. Para Doane (1980) la voz *over* cobra autoridad justamente a partir de su ruptura, de su separación con el espacio diegético: “Es precisamente porque la voz no es localizable, que no puede ser unida, relacionada a un cuerpo, por lo que es capaz de interpretar la imagen, produciendo su verdad” (Doane 1980:168)

Como resumen de esta reflexión podemos decir que en el cine narrativo de ficción, por lo menos en su estilo más clásico, la puesta en escena tiene la función de perpetuar la imagen de unidad sensorial del cuerpo que representa, para lo cual la voz, en cualquiera de sus formas, colabora activamente. Eso es lo que tienen en común todos los tipos de voz. Doane (1980:172) lo resume así:

“..la sincronización ata la voz a un cuerpo en una unidad cuya proximidad sólo puede ser percibida como algo dado; la voz en *off* mantiene el espectáculo de un espacio –extendido pero aún coherente-; el comentario en voz *over* coloca la imagen dotándola con una clara inteligibilidad. En todo esto, lo que debe ser vigilado es una cierta unicidad”.

2.8.2.2 La acusmatización de la voz y su manipulación tecnológica.

El concepto de **acusmatización** (Schaeffer 1966, Doane 1980, Chion 1982 y Rodríguez 1998) que denomina el fenómeno mediante el cual no se ve la fuente de donde proviene aquello que se oye, es decir, disgrega para los sentidos el sonido de su fuente, resulta clave para nuestro estudio. A partir de la creación de la tecnología que permitió grabar el sonido, este concepto cobró una nueva vigencia, sobre todo, para los medios audiovisuales.

Distintos autores (Chion 1982, 1988, 1990b y 1998; Altman 1980,) refieren el carácter mítico y misterioso que la voz acusmatizada ha tenido siempre para las distintas religiones.

Hemos visto que la posibilidad de acusmatizar el sonido en el sentido de separarlo de su fuente, ha hecho posible el sonido cinematográfico como tal. Excepto raras excepciones, como en el *cinema vérité*, desde el estricto punto de vista filmico, el sonido siempre es asincrónico, es decir, producido en un momento distinto al de la banda de imágenes. En ello coinciden la mayoría de los teóricos que han tratado el tema (Percheron; Doane; Altman; Chion; Camper) y es justamente desde esta asincronía entre la producción del sonido y su posterior relación con la diégesis y con la banda de imágenes y, por ende, con el

espectador, que el sonido cinematográfico puede ser analizado, discutido y clasificado como concepto teórico.

Así, entre otras casi infinitas posibilidades en la comunicación audiovisual, la acusmatización posibilita el fenómeno del doblaje. Rodríguez (1998:36) recuerda que esta acusmatización ha permitido: “la creación de nuevos personajes partiendo de una recomposición del par voz – imagen y posibilita mezclar los rasgos físicos y de expresión gestual de un actor con la capacidad de expresión oral y sonora de otro”.

Este proceso de acusmatización como posibilitador del fenómeno del doblaje, lo es en tanto se produce en el terreno de la manipulación técnica del sonido, porque cuando se trata del sonido en la película, a la hora de su emisión - recepción, es justamente el trayecto inverso a la acusmatización. Es la encarnación de las nuevas voces en nuevos cuerpos –lo que Schaeffer (1966) propone llamar “escucha directa” y Chion (1990b:74) prefiere denominar “escucha visualizada”- lo que permite que el ciclo se cierre y la ilusión del doblaje se produzca eficazmente.

En el lenguaje audiovisual, una vez que el receptor reconoce un sonido virtual, es decir, relaciona la forma sonora (que puede haber sido producida por una fuente totalmente ajena a la que expresa) con su experiencia y recuerdo, se configura lo que Rodríguez define como “ente acústico”. Una vez configurado el ente acústico “éste puede ser tratado ya por el narrador como si se tratara de la fuente misma” (Rodríguez 1998:202). Así, las voces dobladoras serían verdaderos entes acústicos asignados a fuentes visuales perfectamente identificables.

Desde el punto de vista de la emisión debemos destacar que el sistema sonoro del cine y todos sus subsistemas son creados y manipulados en función de una artificialidad intencionada de manera que permita conseguir el efecto deseado. Al analizar la tecnología de la voz en el cine, Altman (1985) centra su interés en rebatir la motivación que alegaban los ingenieros de sonido de conseguir mayor naturalidad y realismo como objetivo de la evolución tecnológica de los micrófonos.

A juicio de Altman (1985), el desarrollo y evolución tecnológicos de los distintos procedimientos de manipulación y grabación de las voces en el cine (su texto se refiere específicamente a la evolución de los micrófonos) responden a posibilitar una cada vez mayor eficacia para ofrecer lo que la intencionalidad del director determina y no un sonido pretendidamente “realista o naturalista”.

“Entre la llamada escucha normal, con su selectividad y la selectividad representada por el micro direccional, reposa una función narrativa introducida por la necesidad de *elegir* la dirección en la cual el micro ha de ser activado. Mientras que este modo de escucha direccional quizás se parezca a la escucha binaural natural, difiere de ella en que escucha *por* el espectador.” (Cursivas del autor: Altman 1985:16)

En un trabajo anterior, Altman (1980:79) denuncia las convenciones de la producción del sonido cinematográfico en contraste con la intención que su reunificación con la imagen pretende. Califica así al cine de ventrílocuo porque:

“...es el producto de un esfuerzo para vencer el hueco sonido-imagen, para enmascarar el origen tecnológico del sonido y para permitir al personal de producción (realización) del film expresar su subconsciente, -su barriga- sin miedo a ser descubiertos”.

Como en el apartado anterior señalábamos, la voz es un ingrediente indispensable para la creación del espacio diegético. Gaudreault y Jost (1990:145) llaman “especie de verosímil sonoro” a la tendencia del cine sonoro a privilegiar las voces proscribiendo, por ejemplo, los ruidos demasiado elevados o la música durante los diálogos. El espacio visual imaginario que el film construye requiere de técnicas específicas para espaciar la voz, para localizarla, para darle profundidad y así presentar a los personajes con la consistencia de lo real.

“(En el cine, en efecto, respecto a las condiciones de escucha «*in vivo*», debe reforzarse de manera artificial el nivel y la presencia de la

voz en relación a otros sonidos, para compensar la ausencia de aquellas marcas o señales que, en la escucha directa «binaurale», nos permiten aislar la voz del ruido del ambiente)” (Chion 1982:15-16)

La coloración o el tono sonoro del ambiente, las características de la reverberación y la perspectiva del sonido manifiestan también un deseo de recrear lo que un experto editor de sonido describe como “el bouquet que rodea a las palabras, la presencia de la voz, la manera que combina con el ambiente físico.” (Sturhahn 1974:23) De hecho, incluso, lo que en un film sonoro es decodificado como silencio, está tratado con el sonido ambiente, el color o tono del “silencio” en el espacio que representa.

En este sentido, el mismo Belazs (1952) aboga en su momento porque los cineastas sean capaces de representar esa naturaleza real del sonido; elogia las capacidades de los micrófonos que retienen el color y tono original del sonido que grabaron y los compara con la angulación de la cámara que ofrece una perspectiva propia aunque el espectador siempre mire hacia arriba, hacia la pantalla.

De este modo nos encontramos con que a través de la manipulación técnica, por otra parte, cada vez más desarrollada y eficaz con los actuales sistemas digitales, la participación de las voces en la construcción del espacio sonoro (Rodríguez 1998:227) mediante su planificación y manipulación, junto con las formas de percepción audiovisual, el doblaje se convierte en un elemento nuevo de la obra recreado de manera semejante a los demás subsistemas sonoros de la obra original.

A la hora de definir la experiencia diegética, destacábamos que todos los elementos que componen el lenguaje cinematográfico están al servicio de interpelar al espectador, pero como “individuo incorpóreo”, y embarcarlo así en el viaje inmóvil: las sustancias sonoras del cine y, especialmente la voz, no podían ser menos. En su momento identificamos (Palencia 1998) que el espectador asume el doblaje como un código más, producto de la relación que se establece entre las intencionadas manipulaciones técnicas del sonido y el proceso perceptivo de audiovisión que, desde la recepción, completa y decodifica estos códigos.

2.9 Aspectos perceptivos de la película doblada.

“El cine presenta un espectáculo compuesto de elementos dispares –imágenes, voces, efectos sonoros, música, escritura- cuya puesta en escena, en su sentido más amplio, organiza y apunta al cuerpo del espectador, receptáculo sensorial de los diversos estímulos” (Doan 1985:175)

El debate sobre la preeminencia del sonido o de la imagen en la percepción audiovisual sigue siendo sumamente vivo y un objeto de estudio vigente y dinámico tanto entre los psicólogos de la percepción, los teóricos de la comunicación audiovisual y los estudiosos del cine en particular. Por lo tanto, antes de considerar investigaciones experimentales precedentes a este trabajo, en este apartado incluiremos primero una reflexión en torno a los aspectos perceptivos a considerar por parte del receptor de películas dobladas.

Consideraremos en un primer momento la escucha como fenómeno que afecta directamente al estricto cambio de voces que produce el doblaje y como reflexión en comparación con la distinta recepción que experimenta el espectador de películas subtituladas, alternativa principal al doblaje. Posteriormente consideraremos la audiovisión, como fenómeno específico y posibilitador de la eficacia del doblaje. En cada apartado incluiremos las conclusiones pertinentes obtenidas en la investigación exploratoria que, bajo una metodología cualitativa, llevamos previamente a cabo sobre recepción de doblaje.

2.9.1 Niveles y actitudes de escucha

El modelo de los mecanismos de escucha propuesto por Schaeffer (1966:61-66) y recogido por Rodríguez (1998: 198-207) arroja luz sobre este aspecto de nuestro objeto de estudio.

Según este modelo, existen distintos niveles de escucha, a saber:

1º Oír: es el primer y más simple nivel. Consiste en recibir información a través del oído.

2º Escuchar: supone prestar atención al sonido con el fin de recoger información que interesa. Este segundo nivel puede ser más analítico dependiendo de la voluntad del receptor.

3º Reconocer: supone principalmente el reconocimiento de la fuente sonora o su relación con una fuente sonora que se recuerda.

4º Comprender: supone interpretar en base a la escucha y el reconocimiento. “Comprender es, pues, ir más allá de la identificación de la forma y de la fuente, es producir un nuevo nivel de sentido a partir de la interpretación de aquello que estamos oyendo en función del contexto perceptivo y de nuestra propia experiencia auditiva”. (Rodríguez 1998:203).

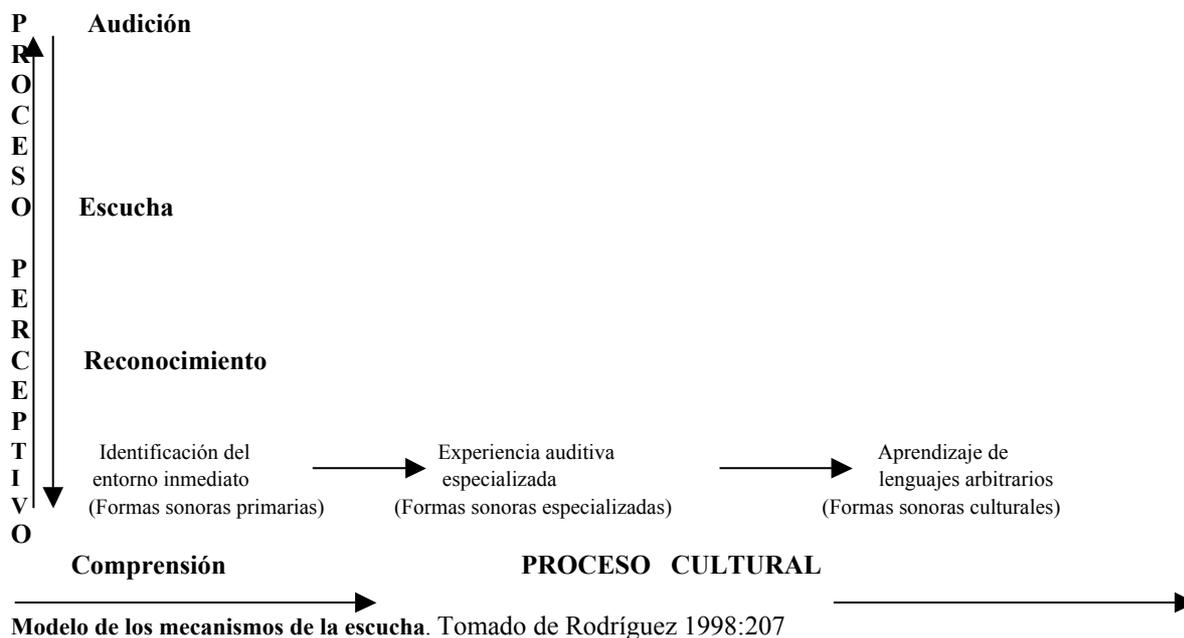
A la par de estos cuatro niveles, es decir, de la asignación de sentido a lo que escuchamos sobre la base de su reconocimiento y percepción consciente, esta comprensión se estructura, a su vez, en tres niveles que descansan sobre la base del aprendizaje.

1º La memoria auditiva del entorno inmediato. Este aprendizaje es adquirido por todo ser humano durante sus primeros años de vida.

2º La experiencia auditiva especializada. Este nivel está referido a la especialización en el reconocimiento y asignación de sentido a las formas sonoras percibidas conforme a los intereses específicos del individuo.

3º El aprendizaje de lenguajes sonoros arbitrarios. Este último y más complejo nivel es el que se refiere a los sistemas sonoros complejos y reglados que exigen códigos de asociación entre formas sonoras y sentido, como es el caso de la lengua.

De manera que la escucha humana descansa sobre dos coordenadas: una perceptiva y otra cultural, según el cuadro que propone Rodríguez:



De acuerdo a este modelo de estructuración de la escucha humana podríamos pensar que desde el eje “proceso perceptivo” los receptores del doblaje oyen, escuchan, reconocen y comprenden, a diferencia de los consumidores de productos subtítulos que desconocen la lengua original que, por lo que a las voces se refiere, solamente alcanzarían los tres primeros niveles.

Por otra parte, desde la coordenada “proceso cultural”, los receptores del doblaje asumen los niveles primero y tercero. Es decir, tras la identificación de las voces en relación a su propio recuerdo, entorno y experiencia, descodifican el sentido de esas voces gracias a su previo conocimiento de la lengua en que éstas se expresan. Y, todavía más allá, podríamos suponer una experimentación del segundo nivel, el de la especialización, en la medida en que el consumo frecuente de doblaje y el interés específico en el consumo cinematográfico, lo favorezcan.

Aunque de manera no experimental, mediante una metodología cualitativa, nuestro trabajo precedente sobre recepción de doblaje confirmó estas hipótesis. Los sujetos entrevistados señalaron como factor de eficacia “la casi simultaneidad entre hablar y oír y la

mayor rapidez en la decodificación de los contenidos verbales mediante la percepción auditiva sobre la percepción visual que implica la lectura de subtítulos.” (Palencia, 1998:149).

La valoración positiva de la escucha de las voces dobladoras por parte de los receptores en dicho trabajo aludía a la percepción más íntegra y reposada de las imágenes en pantalla gracias al mayor tiempo disponible para mirarlas que permite el consumo de doblaje respecto del que ofrece la subtitulación, y a la ausencia de elementos extradiegéticos en la pantalla, como los subtítulos.

Por su parte Chion (1990b:33) propone tres actitudes distintas de escucha:

1ª La escucha causal: servirse de un sonido para informarse de su causa. Este tipo de escucha es la más frecuente y también la más susceptible de verse influida o engañada.

2ª La escucha semántica: está referida a interpretar un mensaje gracias a un código y un lenguaje, como la lengua.

3ª La escucha reducida: consiste en atender a las formas del sonido independientemente de su causa y su sentido. Esta actitud es infrecuente.

Podríamos pensar que los receptores de doblaje adoptan las dos primeras actitudes, mientras que los receptores de productos subtitulados que desconocen la lengua original, respecto a las voces originales ejercitan la primera y la tercera. Es decir, son capaces de reconocer las causas de esas voces, sus fuentes respectivas, pero no de interpretar su lengua. En cambio, el sonido de unas voces cuyo mensaje no es posible de ser interpretado, animaría el difícil ejercicio de la escucha reducida.

En la investigación exploratoria previa a la presente pudimos concluir que “el conocimiento de la lengua destino no sólo es la razón del doblaje, sino que es un factor determinante en la percepción de eficacia de este proceso de transferencia lingüística desde la recepción. La inteligibilidad de los contenidos verbales del doblaje estructura la visión de la

interpretación de los actores en pantalla y eleva el nivel de escucha del receptor en la medida en que es capaz de identificar la causa de las voces que escucha, reconocerlas y, además, descodificar eficazmente su sentido” (Palencia 1998:151).

2. 9.2 La audiovisión

Durante más de cien años de vida del cinematógrafo, sus teóricos y sus críticos han debatido sin aparente unanimidad cuál de las dos sustancias principales de las que se compone el cine: sonidos o imágenes, es la principal o preeminente. Históricamente, en una primera instancia, como hemos visto, tanto los devotos como los detractores del invento echaron en falta su carencia de sonido sincrónico.

Más tarde, la aparición del sonoro, llamado entonces con inconsciente propiedad *talkies* o *cinéma parlant*, encontró también sus detractores. Para algunos la crítica se fundamentaba en la pérdida de lo que consideraban la poesía de la gestualidad silente en la que entendían que residía el arte del cinematógrafo. Otros, rechazaban el carácter teatral que fue adquiriendo el invento al centrar el uso del sonido en la palabra sincrónica y, con ello, el retraso en la investigación y práctica de las casi infinitas posibilidades expresivas tanto del sonido como de la imagen cinematográfica.

Desde los primeros tiempos del sonoro, junto a consistentes teorías que reclamaban la preeminencia de la imagen y, por ende, la subordinación del sonido, en un intento de avanzar más allá del sincronismo teatral, encontramos otras, que con autoridad semejante estudiaron las ricas posibilidades que el sonido cinematográfico ofrecería como estructurador y modificador de las cualidades significantes de esa misma imagen (Eisenstein, Pudovkin, Kulechov, Vertov, entre otros).

Según Altman (1980 a:14) el debate se ha centrado en dos errores. El primero, al que llama “la gran falacia histórica” que considera que los filmes sonoros están compuestos de dos fenómenos paralelos y simultáneos: imagen y sonido. A su juicio, esta primera falacia conlleva la jerarquización cronológica de ambos componentes. Irónicamente dice: “el sonido

fue añadido a la imagen: ergo, en el análisis del cine sonoro debe ser tratado como un suplemento que la imagen es libre de tomar o dejar, como elija”.

La segunda, sería la “falacia ontológica” que llevaría a considerar que el cine es un medio visual y las imágenes, las principales portadoras de estructura del film. En este sentido arremete contra teóricos como Arnheim en su “*New Lacoön*” o Kracauer que en su *Theory of Film* asevera: “Las películas sonoras cumplen (son coherentes) con el espíritu del medio sólo si lo visual toma la dirección” (“*Films with sound live up to the spirit of the medium only if the visuals take the lead in them*”: citado por Altman 1980:14) O contra Bettetini que en su obra *The Language and Thechnique of the Film*, asegura que “la esencia del cine es básicamente visual, y toda la intervención sonora debe limitarse a sí misma a un acto necesario y justificado de integración expresiva” (citado por Altman ibidem).

Durante el recorrido teórico precedente, hemos visto que el sonido cinematográfico no puede ser analizado, categorizado, teorizado y decodificado, más que a partir de su relación con la imagen a la que acompaña. El principio de la no sincronización o asincronía que estuvo de moda en los inicios de la era sonora, condenaba el sonido sincrónico (*on*) como redundante y se basaba justamente en privilegiar el sonido *off* (Percheron: 1980:16). Opuesto a esta tendencia, el cine clásico de Hollywood centraba su atención en el hablante durante las escenas dialogadas y en el sonido sincrónico, redundante, práctica aún mayoritaria en el cine narrativo de ficción.

Creemos, junto con cada vez más autores, que más allá de la escucha, más allá de ver y oír, el espectador de cine, experimenta una audiovisión. De ahí que Chion proponga que se le llame “audio-espectador”.

Como acertadamente opina Altman (1985:68):

“Imagen y sonido: en la medida en que cada uno refuerza la mentira del otro, no dudamos en creer en ambos. Esta ideología complementaria, dado que cada banda ayuda a esconder el trabajo de la

otra, debe necesariamente desmentir, abortar, cualquier teoría que postule la primacía de cualquiera de las dos bandas”.

Igualmente acertada y complementariamente Doan (1985) reclama:

“El punto no es que “necesitemos” términos con los cuales describir, honrar y reconocer la autonomía de un material sensible particular, sino que debemos atrevernos a pensar (acerca de) la heterogeneidad del cine”. (Doan 1985:166)

Entendemos así que en el lenguaje audiovisual el sonido y la imagen se influyen mutuamente gracias al **valor añadido** recíproco que Chion (1990b:16) define como:

“el valor expresivo e informativo con el que un sonido enriquece una imagen dada, hasta hacer creer, en la impresión inmediata que de ella se tiene o el recuerdo que de ella se conserva, que esta información o esta expresión se desprende de modo <natural> de lo que se ve, y está ya contenida en la sola imagen”.

Este fenómeno funciona especialmente en el marco del sincronismo sonido/imagen y de manera recíproca. Es decir, gracias a esta conjunción sonido – imagen, la percepción es nueva y distinta que la percepción auditiva y visual de ese mismo sonido o esa misma imagen separadamente.

“...el sonido hace ver la imagen de modo diferente a lo que esta imagen muestra sin él, la imagen, por su parte, hace oír el sonido de modo distinto a como éste resonaría en la oscuridad” (Chion 1990b:31)

Bela Balazs (1952:117) lo expresa muy gráficamente cuando dice:

“Así como la sombra y el valor de un color cambia de acuerdo con los otros colores cercanos a él en una pintura, así el timbre

de un sonido cambia en concordancia con la fisonomía o el gesto de la fuente visible del sonido, vista con el sonido mismo en un film sonoro en el cual las impresiones acústicas y ópticas son ligadas equivalentemente dentro de una única imagen (*picture*)". (Balazs 1952: 117)

Cuando se trata de un texto oral, el sonido de ese texto, no es sólo una reiteración o una negación de la imagen. Así lo demuestra Chris Marker en su documental *Lettre de Sibérie* (1957) en el que una misma secuencia con tres textos distintos en *off* que explican sentidos opuestos desde perspectivas políticas antagónicas: estalinista, antiestalinista y una tercera pretendidamente neutra y descriptiva. Bajo cada uno de los distintos textos en *off*, la imagen cobra significados distintos.

Estamos así de acuerdo con Chion (1990 b:19) en que el valor añadido del texto sobre la imagen es algo más que una opinión o una tautología, el sonido del texto y, más concretamente la voz, estructura la visión "enmarcándola rigurosamente".

En el caso de la voz, Eisler y Adorno (citados por Doan 1985:171) hicieron notar que el espectador está siempre al tanto de la divergencia entre el cuerpo representado y su voz, porque la voz grabada posee una profundidad que contrasta con la chatedad de la imagen bidimensional de la pantalla. Para estos autores esto explica parcialmente la función de la música en el cine: primero, en los filmes mudos, para ocultar el ruido del proyector; luego, en las tomas habladas, para disimular el hueco entre la voz y el cuerpo. Sin embargo, Doan (1985:171) considera que "la subordinación de la voz a la pantalla como el sitio del despliegue del espectáculo hacen que la visión y la escucha trabajen juntos en la manufactura de la "alucinación" de un completo mundo sensorial".

En esta misma dirección Rodríguez (1998:220) rebate uno de los argumentos que pretenden dar fundamento a la extendida creencia de que existe una preponderancia del sentido de la vista sobre el del oído, como es la convicción de que el sistema auditivo se rige por el sistema visual. Afirma que, mientras la vista cubre 180 grados del entorno y depende de la luz y la posición del receptor, el oído cubre los 360 grados y no depende de ninguno de esos factores. El oído, así, indica a la vista el origen de los sonidos. Rodríguez concluye: "La

fenomenología del doblaje demuestra, también, lo fácil que es engañar desde el oído al sentido de la vista”.

Altman (1980: 73) discurre por un camino ¿inverso o semejante? cuando asegura que “lejos de ser redundante, el sonido (cinematográfico) tiene una cualidad enigmática fundamental que confiere sobre la imagen la cualidad de una respuesta (...) La imagen, en términos de sonido, siempre tiene la naturaleza básica de una pregunta”.

Analícemos: para Rodríguez el sonido alerta al perceptor de que algo suena, sin necesidad de estar delante de la fuente. Para Altman el sonido cinematográfico es la respuesta a la pregunta ¿qué o cómo suena la imagen (posible fuente de sonido) que tenemos delante? Sin embargo, para Doane (1980) el sonido cinematográfico, especialmente la voz, exige, demanda, su encarnación en un cuerpo, en este sentido, el sonido es la pregunta ¿qué o quién suena o habla? y la imagen su respuesta. Igualmente, para Bordwell y Thompson (1985; 1995) “el sonido indica qué mirar”. Como vemos, el juego de reciprocidades puede ser infinito.

Respecto a este juego de intereses e influencias entre sonido e imagen, oído y vista, Bailblé (1989:115) dice:

“Mientras que la vista elige su dominio, mientras que la mirada alcanza las lejanías, el oído padece casi pasivamente las energías desemejantes, el estallido o la debilidad de los sonidos circundantes. Inversamente, y ayudado en ello por las leyes de la propagación, descubre y oye a través, incluso en el momento en que la vista, direccional, cegada por los obstáculos, se deja sorprender fácilmente”.

Esta cita de Bailblé nos permite entender la relación sonido-imagen cinematográficos. La imposibilidad de eludir la escucha posibilita al espectador cinematográfico, que elige mirar la pantalla, decodificar los sonidos cinematográficos, de cualquier tipo, en relación a la imagen que frecuentemente le sorprende. En este contexto cobran significado incluso los sonidos *off*.

Respecto a nuestro objeto de estudio, podemos concluir que el receptor de películas dobladas ha elegido el dominio de su vista, es decir, ha optado por permanecer atento a la pantalla y dejar que el emisor decida sobre el oído de su audiencia, que, sentada en la sala de proyecciones, no puede “abandonar” la escucha mediante el alejamiento.

“El camello-sonido sigue pasando por el ojo de la aguja visual. (...) El sonido sigue siendo lo que nos hace ver en la pantalla lo que él quiere que veamos en ella” (Chion 1990b:137-138).

2.9.2.1 La sincronía como base de la audiovisión en el doblaje

La sincronía entre sonido e imagen es fundamental para estimular la eficacia de esta percepción audiovisual. Tal y como afirma Rodríguez (1998:170) a la hora de analizar la forma como el ser humano utiliza la sincronía para coordinar los sentidos de vista y oído:

“La coincidencia en el tiempo entre las variaciones de los estímulos visuales generados por cualquier objeto físico y los cambios acústicos de un sonido, estimula una asociación perceptiva inmediata entre ese sonido y el citado objeto, desencadenando en el receptor una fuerte sensación de que el objeto que ve es la fuente física del sonido que está escuchando”.

Para Fodor (1976: 40-60) los factores que influyen en la mayor o menor detección de asincronía fonética del doblaje por parte de los receptores se resumen en:

- 1) La existencia de otras fuentes sonoras, especialmente ruidos como disparos o explosiones: el espectador presta menos atención a lo que dicen los actores.
- 2) La deficiente iluminación disminuye la capacidad observación de los movimientos articulatorios.

3) El tamaño de la pantalla y el tipo de plano y de ángulo de la cámara. Siendo más susceptibles de atención la pantalla más grande, el plano más corto y el ángulo más frontal.

4) La acción de la secuencia puede desviar la atención del espectador fuera del discurso y los movimientos articulatorios.

5) La relación entre los sistemas vocálicos de la lengua origen y la lengua término.

6) Las características del espectador:

a) Tipo de atención subjetivo y objetivo. Los hombres, mayormente proclives a la observación de detalles (objetivo) perciben con más facilidad el asincronismo que las mujeres más inclinadas a fijar la atención en el conjunto global de estímulos (subjetivo).

b) La edad. Los niños y adolescentes tienen mayor capacidad de percepción de los movimientos faciales y gestos corporales.

c) Inteligencia y cultura.

d) Profesión. Aquellos profesionales más acostumbrados a fijar su atención en la expresión de las personas con las que trabajan o con más consciencia de su propia manera de expresarse tendrán más sensibilidad en la relación entre pantomima y sonidos. Ejemplo: actores, profesores, abogados, trabajadores en empresas ruidosas.

Entendemos así que la influencia recíproca entre sonido e imagen en su sincronía se constituye como uno de los factores con más peso específico para la valoración gratificante del doblaje por parte de sus receptores. Su percepción de eficacia está determinada tanto por las características perceptivas de la audiovisión, como por la asignación del valor que el receptor deposita en el doblaje como estimulante en la asunción de la experiencia diegética,

mediante una actitud de complicidad frente a la pantalla. Este tipo de recepción “cómplice” resulta a su vez favorecida o no por las propias características culturales, sociales y perceptivas del sujeto.

“El cine es a fin de cuentas, una fábrica de ilusiones. Digamos entonces que el doblaje busca dar una ilusión de una ilusión”. Caillé (1960:108)

Desde la reflexión consciente de la experiencia perceptiva del doblaje, los sujetos entrevistados en nuestro trabajo precedente consideraron como satisfactoria la relación entre el referente visual y su expresión fonoestésica. Pudimos también concluir que “una vez integradas en la banda sonora del filme, las voces dobladoras se comportan como verdaderos entes acústicos del lenguaje audiovisual en las películas dobladas de modo que los receptores identifican su causa en los actores en pantalla” (Palencia 1998:149).

Sin embargo, en dicho trabajo conseguimos identificar, desde el punto de vista consciente de la recepción, como factores ineficaces del doblaje:

- el conocimiento por parte del receptor de las dos lenguas involucradas en el doblaje: la original y la de destino.
- la ruptura del binomio ideal: una voz – un actor.
- la identificación de una voz con su referente real que no se corresponde con la imagen en pantalla.

Hemos debatido el primer factor en el apartado pertinente, pero los dos últimos factores de ineficacia nos llevan a pensar en una especie de asincronía, no ya física o articuladora, sino una distinta que se produce en la mente del receptor en la medida en que le exige una complicidad más allá de la que está dispuesto a aportar a priori cuando decide ver una película doblada y, en este sentido, subordinar los códigos y servidumbres del doblaje a su motivación prioritaria de fruición de la experiencia diegética. De lo que se trata ahora, es de identificar de manera experimental, si el doblaje modifica y en qué medida, la credibilidad-verosimilitud de sus personajes y su carácter o actitud.

2.10 Investigaciones experimentales precedentes

El presente estudio se enmarca dentro de la línea de investigación “Modelo Acústico de Credibilidad (MAC) de la voz en el marco de la gramática de la expresión fonostésica”, PB93-0850 de la DGICYT en la que, bajo la dirección del doctor Emili Prado Pico, pudimos participar junto con un grupo de investigadores del Departamento de Comunicación Audiovisual de la UAB entre 1994 y 1997.

El MAC es, a su vez, la continuación de una línea de investigación abierta en el Departamento de Comunicación Audiovisual y Publicidad de la Universidad Autónoma de Barcelona en 1979 bajo la dirección de Emili Prado, cuyos primeros frutos se recogen en varias tesis de licenciatura y cristalizan en una primera aportación sustancial en la tesis doctoral “Imágenes Auditivas en la Radio” (1987) realizada por el doctor Armand Balsebre.

En ese mismo año, y también en el seno del Grup de Imatge, So i Sintesi (GRISS) que dirige el doctor Prado, se elabora el proyecto de investigación “La gramática de la expresión fonostésica y la representación imaginativo – visual en los sistemas integrales de percepción de la voz” que obtiene una ayuda del Vicerrectorado de Investigación de la UAB y posteriormente es aprobado como proyecto de la DGICYT PB88-0236 en la convocatoria de 1988.

En el marco de este proyecto se desarrollan las tesis doctorales de Ángel Rodríguez “La construcción de una voz radiofónica” presentada en noviembre de 1989; la de José Luis Terrón “El silencio en el lenguaje radiofónico”, presentada en enero de 1992; la de Juan José Perona “El ritmo en la Expresión Radiofónica” presentada en diciembre de 1992 y la de María Teresa Soto “Influencia de la percepción visual del rostro del hablante en la credibilidad de su voz”, presentada 2000, así como el trabajo de investigación, realizado por esta autora, “Factores de eficacia del doblaje en el cine narrativo: un estudio desde la recepción”, con el que obtuvimos el grado de suficiencia investigadora en septiembre de 1998. Los resultados de estos trabajos constituyen un corpus conceptual clave para el presente estudio y seguidamente son glosados.

2.10. 1 El Modelo Acústico de Credibilidad (MAC) de la voz

El MAC se proponía profundizar en el conocimiento del funcionamiento de la voz en los diferentes contextos comunicativos, a la luz de las aportaciones realizadas en una investigación precedente: “Gramática de la expresión fonostésica y la representación imaginativo-visual en los sistemas integrales de percepción de la voz.”

Mediante un corpus sonoro compuesto por locuciones efectuadas por profesionales y no profesionales de ambos sexos sobre textos portadores controlados, se aplicaron pruebas de percepción basadas en el diferencial semántico de Osgood y, por otra parte, se practicó el análisis acústico de las voces del corpus.

El objetivo era determinar la credibilidad, que los sujetos experimentales atribuían a cada una de las voces del corpus mediante dos tipos de percepciones: auditiva y audiovisual

Esta investigación consiguió confirmar (Prado, E. et al: 1997:2-3) que:

- Las características formales del texto portador influyen en el grado de credibilidad, siendo más creíbles los textos formales que los informales tanto en la percepción auditiva como en la audiovisual.

- La capacidad de los locutores profesionales de generar coherentemente una variedad de su voz, predomina en la atribución de credibilidad por encima de las características formales del texto portador, siendo en ambas percepciones, más creíbles las locuciones efectuadas por profesionales.

- El sexo del locutor no influye en la atribución de credibilidad a su voz de forma discriminante. Tampoco en la relación entre las variables sexo del locutor y sexo del sujeto experimental.

- La percepción visual del hablante influye en la atribución de credibilidad que se le otorga sin eliminar el peso de las variables acústicas, pero reduciendo su peso global. Es decir, la atribución de credibilidad es semejante en la percepción auditiva como en la audiovisual, pero el peso de las principales variables que explican la credibilidad, varía en relación al tipo de percepción. Lo que confirma un grado de influencia reducido de la percepción visual del hablante.

- Sobre la estructura lingüística de toda comunicación oral se encabalga un sistema de niveles expresivos específicamente acústicos que se articulan formando la expresión fonostésica. Así se encontró que una frecuencia fundamental superior a los 145 Hz de los profesionales hombres, reduce el índice de credibilidad atribuida. En el caso de las mujeres profesionales se produce la misma lógica sólo que la influencia del texto portador es más manifiesta, situándose el punto de inflexión en los 181Hz.

2.10.2 La expresión fonostésica.

El concepto de “**expresión fonostésica**” es adoptado por el GRISS, tomado de Ivan Fónagy (1983) para avanzar en el ambicioso objetivo de progresar en la formalización de una gramática de la expresión oral en los sistemas integrales de percepción de la voz propuesta en el proyecto de investigación “La gramática de la expresión fonostésica y la representación imaginativo-visual en los sistemas integrales de percepción de la voz” (Prado et al. 1989).

En el marco de este proyecto se desarrolla la tesis doctoral de Ángel Rodríguez (1989) en la que estudia los rasgos sonoros de la comunicación oral que se encabalgan en las estructuras léxicas y gramaticales, rasgos que, sin provenir de ellas, las matizan y modifican sus contenidos generando sensaciones y emociones al receptor.

Para esta característica del lenguaje oral recoge el concepto de “expresión fonostésica” y lo define como:

“La expresividad sonora transmitida mediante los rasgos de la voz que comunican acústicamente información sobre el gesto, la actitud, el carácter, el aspecto físico y el contexto del emisor; o bien sobre la forma, el tamaño, el color, la textura, el tipo de movimiento, etc., de aquello que describe oralmente el emisor. Estos rasgos sonoros no pertenecen a las formas léxicas ni a las formas gramaticales ni tampoco son útiles para identificar fonemas o diferenciar unas estructuras lingüísticas de otras” (Rodríguez 1989:39)

De manera experimental Rodríguez (1989:254-255) consigue demostrar:

- Que el sonido de cada voz lleva asociada una imagen que se repite en la mayoría de los oyentes sin que ésta tenga que ver necesariamente con su imagen física real.

- Que la manipulación de los aspectos formales del sonido de la voz permiten conseguir por parte del receptor una imagen del hablante predeterminada y que puede ser completamente distinta, física y emocionalmente a la del emisor.

- La existencia de una expresión autoacústica (constituida por el conjunto de rasgos de carácter fonoestésico que aportan al oyente información general sobre el emisor), especialmente aquel tipo de expresión que informa respecto a su aspecto físico (idiográfica) y a su carácter o actitud (caracterial o afectiva).

Esta capacidad demostrada de crear distintos tipos de personajes mediante la coherente manipulación de los elementos acústicos de la voz, resulta especialmente relevante para los objetivos de nuestro estudio. A partir del concepto de **expresión autoacústica**, como capacidad sonora de las voces para aportar información sobre el carácter de sus emisores, fundamentamos el corpus de la fase experimental del presente trabajo, como más adelante explicamos en el apartado pertinente.

2.10.3 El silencio en el lenguaje oral

Dentro del mismo proyecto, la tesis doctoral de J.L. Terrón (1991) realiza un exhaustivo análisis teórico sobre el silencio. Aunque referido al lenguaje radiofónico, el trabajo es una profundización sobre el valor del silencio como rasgo expresivo dentro de la comunicación oral cuyas conclusiones consideramos clave para nuestro objeto de estudio.

En “El silencio en el lenguaje radiofónico”, Terrón (1991), tras definir escuetamente el silencio como “lo no audible”, lo sitúa bajo dos coordenadas principales: como un fenómeno relativo (el silencio absoluto no existe) y como un fenómeno perceptivo (el silencio existe sólo en relación con el sujeto que lo percibe). Y demuestra que el silencio es una pauta significativa más dentro de todo proceso de comunicación.

Terrón concluye que la significación del silencio dependerá tanto del contexto sociocultural como del contexto discursivo. De modo que en cada cultura, dependiendo de sus propios valores, por un lado, y de las formas y usos de cada lengua o habla, por otro, el silencio tendrá valores y significados específicos que pueden ser distintos en contextos lingüísticos y culturales diferentes.

“El silencio integrado en el lenguaje tiene la capacidad de significar, entre otras razones, porque existen un significado, un significante y un intérprete. La relación entre silencio y significado, si no hay ninguna patología, está convencionalizada” (Terrón 1991: 638)

A pesar de que el interés de su investigación se centra en el lenguaje radiofónico, Terrón (1991:635) afirma también que el silencio puede expresarse con o sin claves visuales. Es decir, puede darse un silencio verbal mientras que la expresión quinésica del emisor puede estar emitiendo un mensaje.

Consideramos que lo anterior, así como la siguiente conclusión de Terrón respecto al valor del silencio como integrador del lenguaje verbal, resulta relevante para el fenómeno del doblaje:

“El silencio se sitúa fuera del lenguaje en tanto que permite su construcción y, a la vez, es capaz de situarse en su interior como una expresión más.

La relación entre silencio y palabra no hay que plantearla como antagónica. Merced a los silencios las palabras se construyen y se delimitan. Del mismo modo, también podemos considerar a los silencios como una palabra más” (Terrón 1991:635)

Para el objeto que nos ocupa, consideramos que las exigencias de la sincronía articuladora, por un lado y la transferencia cultural que toda transferencia lingüística implica, por otro, pueden modificar este papel del silencio en el producto doblado.

Es decir, durante el doblaje, las pausas y los silencios se modifican, tanto en su duración como en su situación respecto del modelo sintáctico de las dos distintas lenguas implicadas, la lengua fuente y la lengua destino. Esta modificación puede hacer que la significación de cada pausa y silencio cobre un valor distinto en el discurso doblado que en el original.

Terrón (1991:636) afirma que “... una manera de no caer en una decodificación aberrante (del silencio) es relacionar solidariamente al silencio con el resto de la enunciación.” Esta relación solidaria puede resultar modificada por las exigencias sincrónicas del doblaje. Lo relevante será identificar si la decodificación resulta aberrante o puede, por el contrario, merced al anclaje de la voz en un cuerpo, es decir, merced a su conversión en un ente acústico consistente, junto con el contexto discursivo en el que se expresa, ser decodificado correctamente.

2.10.4 El ritmo en la expresión oral

En el seno del mismo proyecto se desarrolla la tesis doctoral de J. J. Perona (1992) sobre el ritmo en la expresión radiofónica que aporta conceptos útiles para nuestra investigación. Perona (1992:39) define el ritmo como:

“El resultado de una actividad perceptiva capaz de captar una estructura determinada por una sucesión de fenómenos, aislados o asociados, que se repiten regular o irregularmente en el tiempo”.

Afirma que el lenguaje radiofónico es un “*continuum*” en el que intervienen distintas estructuras rítmicas superpuestas y estrechamente interrelacionadas, pero también jerárquicas, a saber:

- a. Estructura rítmica interna
- b. Estructura rítmica melódica.
- c. Estructura rítmica armónica
- d. Estructura rítmica patronal
- e. Estructura rítmica global

Para el objeto que nos ocupa, las tres primeras estructuras arriba mencionadas resultan claves. De acuerdo con Perona, la estructura rítmica estará definida por el modo en que se sucedan las unidades mínimas en cada uno de los sistemas radiofónicos.

Referidas al lenguaje hablado, uno de las cuatro sistemas que componen el lenguaje radiofónico (verbal, musical, silencio y efectos sonoros) la unidad rítmica mínima lo constituye la sílaba. Su número y duración definirán la estructura rítmica interna del discurso oral.

Cuando se toman como referentes unidades lingüísticas más amplias, como el grupo fónico (“parte mínima del enunciado dotada de significado diferenciado y situada entre dos pausas” (Perona 1996:203) la variación de la estructura rítmica interna, es decir, la modificación del número en el tiempo y la duración de cada grupo fónico y, dentro de él, de cada sílaba, puede modificar el sentido mismo del enunciado.

Esto, dice Perona (1996:203), “le confiere a la estructura rítmica interna un poder extraordinario, hasta el punto de que según la forma que ésta adopte, un mismo contenido semántico podrá ver sustancialmente alterado su significado”.

La estructura rítmica melódica estará definida por la entonación con que se pronuncien los grupos fónicos. En castellano las unidades melódicas suelen coincidir con los grupos fónicos, aunque esta coincidencia puede verse afectada por la intencionalidad.

La estructura rítmica armónica nace de la combinación de dos o más sistemas que conforman el lenguaje radiofónico o de la combinación de dos fuentes sonoras de un mismo sistema (dos o más voces).

“La estructura armónica es, por tanto, fruto de una conjunción ordenada de sustancias susceptibles de repetirse regular o irregularmente en el tiempo. Su ritmo vendrá marcado por la duración y la frecuencia de repetición de esas sustancias y por los ritmos propios de las mismas. Pero además, en tanto que combinación de sonidos, también lo será de timbres, tonos e intensidades diversas.” (Perona 1996:205)

Según las conclusiones de Perona (1992:204) las dos estructuras rítmicas: interna y melódica, referidas al lenguaje verbal, son capaces de informar sobre elementos emocionales en el contexto del discurso, mientras que la estructura rítmica armónica surge en la combinación de dos o más sistemas (verbal, musical, silencio y efectos sonoros).

Estas tres estructuras aplicadas al lenguaje oral inevitablemente se verán afectadas con el doblaje. No sólo por la distinta estructura rítmica interna y melódica inherente a cada lengua, sino por la variación dentro de las tres estructuras rítmicas a que el discurso se verá sometido por la exigencia del sincronismo articulatorio del doblaje. Si consideramos la obra audiovisual en su conjunto igualmente es de suponer que la estructura rítmica global se verá afectada.

2.10.5 La influencia de la percepción visual del rostro del hablante

La tesis doctoral de María Teresa Soto (2000), “Influencia de la percepción visual del rostro del hablante en la credibilidad de su voz”, desarrollada en el marco del proyecto “Modelo acústico de la credibilidad (MAC) de la voz en el marco de la gramática de la expresión fonostésica” (Prado et al 1997), resulta especialmente relevante para el presente trabajo de investigación. Es por ello que nos detendremos en sus resultados y, fundamentalmente, en el interesante y exhaustivo debate sobre los distintos modos de percepción del habla que su marco teórico plantea.

Soto (2000) lleva a cabo un experimento cuyo objetivo es detectar la influencia de la percepción visual del rostro del hablante en la atribución de credibilidad. Para ello construye un corpus audiovisual consistente en dos textos distintos, uno informativo (denominado formal) locutado frente a cámara por locutores profesionales y no profesionales, hombres y mujeres y otro texto narrativo (llamado informal), igualmente locutado por locutores profesionales y no profesionales, hombres y mujeres. El mismo texto, con ausencia de imagen constituyó el corpus sonoro del experimento.

Ambos corpus, audiovisual y sonoro, son expuestos de manera separada a la consideración de 16 grupos experimentales de audiencias cautivas constituidos por 20 sujetos en una proporción equivalente de hombres y mujeres, 320 sujetos en total. Es decir, 8 grupos experimentales (160 sujetos) audiovisitaron el corpus audiovisual y 8 grupos experimentales (160 sujetos) solamente escucharon el corpus sonoro.

Mediante el diferencial semántico de Osgood, los sujetos experimentales valoraron la credibilidad de las voces en escalas de 1 a 7 grados y en función de 12 pares de adjetivos bipolares, validados previamente como delimitadores de credibilidad en su interacción, más la variable credibilidad, también como adjetivo bipolar.

En un primer momento, mediante el análisis estadístico descriptivo, Soto consigue demostrar que la percepción visual del rostro del hablante “no determina la credibilidad de sus

voces; que el peso en la atribución de credibilidad depende firmemente de los factores sonoros de la producción del habla” (Soto 2000:397).

Asimismo, encuentra que la percepción de la imagen no afecta por sí sola a la credibilidad de las voces:

“Los sujetos experimentales no varían estadísticamente sus impresiones únicamente por ver las imágenes de los locutores. El peso en la atribución de credibilidad dentro del sistema voz-imagen descansa, en líneas generales, en las voces” (Soto 2000:184).

Sin embargo, en un segundo nivel de profundidad en el análisis estadístico inferencial, sus investigaciones la llevaron a concluir que la percepción visual del rostro del hablante sirve como acicate consistente a la hora de precisar los juicios.

Así concluye que:

“(…) las imágenes ratifican y perfilan los grados de credibilidad de las voces de los hablantes y brindan seguridad a los perceptores en las tareas de enjuiciamiento”. (Soto 2000:398).

Soto demuestra así, de manera experimental, que “el sonido estructura la visión” (Chion:1990b, Rodríguez:1998) y, al mismo tiempo, constata, el “valor añadido” con el que se enriquecen mutuamente, imagen y sonido en la percepción audiovisual.

El trabajo de Soto es prolijo en conclusiones precisas sobre el grado y la influencia de las variables consideradas, tales como sexo del sujeto, sexo del locutor, tipo de texto, frecuencia fundamental de las voces, velocidad del habla, así como de las variables que definen la credibilidad, bajo cada uno de los modos de percepción: auditivo o audiovisual.

Aunque, como apuntábamos en el apartado correspondiente, el concepto “credibilidad” en el trabajo de Soto está referido a la dimensión informativa, dentro del carácter multidimensional del término, consideramos que sus conclusiones igualmente arrojan luz sobre nuestro trabajo en la medida en que la percepción bimodal es también un fenómeno

clave para nuestro objeto de estudio. Dado el carácter extremadamente exhaustivo de sus conclusiones, solamente destacaremos aquí aquellas que nos parecen relevantes para nuestros objetivos. A saber:

Los perceptores de la modalidad sonora, atribuyen credibilidad en base a la variable verosimilitud; en cambio, los que audiovisonan el habla, tienden a atribuirla a partir de la variable inteligencia.

Variabes como verosimilitud y seguridad son un aporte de la voz al modelo de la credibilidad audiovisual, pero mientras la verosimilitud pierde peso en la modalidad audiovisual, la seguridad lo gana.

Las voces que locutan textos formales (aquellos de carácter más informativo) obtienen mayores cuotas de credibilidad, que las que locutan textos informales (aquellos que consideramos más cercanos a un texto interpretado de ficción). Pero la percepción audiovisual provoca una mayor diversidad en la evaluación de los textos informales. Es decir, la audiovisión produce que los sujetos evalúen con mayor precisión la credibilidad de textos informales.

La profesionalidad del locutor influye también en la credibilidad del hablante. Los profesionales son más creíbles que los no profesionales independientemente de la modalidad de percepción, auditiva o audiovisual. Sin embargo, como sucede con el texto, la percepción de las imágenes de los hablantes permite una mayor precisión en la evaluación.

De estas dos conclusiones, Soto deduce que “independientemente de la modalidad de percepción, la profesionalidad en primer lugar, y el tipo de texto en segundo, afectan a la credibilidad de la voz de los hablantes (...) El peso de las variables acústicas prima sobre las visuales”. (Soto 2000:403)

Del cruce estadístico entre modalidades de percepción y sexo de los locutores, Soto (2000:405) concluye, entre otras consideraciones que, mientras la variable inteligencia, descansa en la voz de los locutores masculinos y en la imagen de los femeninos, la percepción de verosimilitud descansa en las voces de ambos.

En este punto nos detenemos para relacionar las conclusiones arriba anotadas con nuestro objeto de estudio. Consideremos tres conceptos que están involucrados en el fenómeno del doblaje: verosimilitud, profesionalidad y tipo de texto. Deducimos de las pruebas experimentales de Soto al respecto, que si las variables profesionalidad y tipo de texto resultan sustanciales en la atribución de credibilidad de los hablantes, podemos pensar que el trabajo de los dobladores, siendo como es un trabajo profesional, ejecutado por actores y actrices que poseen, por el aprendizaje y la experiencia, un fuerte dominio y control sobre sus voces, que además interpretan textos de ficción, es decir, más cercanos a lo que el trabajo de Soto describe como informales, y si además, la variable verosimilitud descansa fundamentalmente en las voces (en las sustancias sonoras) independientemente del sexo del hablante, podemos pensar, decíamos, que el doblaje puede mantener para el perceptor valores coherentes e incluso superiores de la variable credibilidad (bajo su acepción de verosímil) respecto de las voces originales de las películas dobladas.

Respecto a variables acústicas de la voz explotadas, tales como la frecuencia fundamental de las voces o la velocidad del habla, Soto concluye:

Sobre frecuencia fundamental: “las voces con frecuencias medias son significativamente más creíbles que las agudas y éstas a su vez, significativamente más creíbles que las graves” (Soto 2000:413). Sin embargo, la percepción audiovisual de los hablantes afecta a estos juicios; en mayor proporción a los que poseen voces con frecuencias más extremas: graves y agudas. “Las voces graves que son percibidas audiovisualmente tienen más probabilidades de resultar no creíbles y las voces medias y agudas más probabilidades de resultar creíbles”. (Soto 2000:414)

Sobre velocidad del habla: Los locutores que leen con velocidad más rápida son significativamente más creíbles que los que leen a velocidad lenta, bajo cualquiera de las dos modalidades de percepción. “A medida que desciende la velocidad del habla, descienden los grados de credibilidad” (Soto 2000:415).

En nuestro trabajo, los aspectos referidos a las características acústicas de las voces, semejantes a los estudiados por Soto (2000), el objetivo consiste en identificar en qué medida la variación de la frecuencia fundamental de las voces doblada y original respectivamente, afecta a la coherencia en la atribución de la credibilidad y de los respectivos caracteres en cada uno de los actores y actrices participantes. Respecto a velocidad del habla, creemos que si bien es un aspecto que puede verse modificado por el doblaje, tendrá una incidencia menor debido a las exigencias de la sincronía articulatoria a la que se ve sometido el fenómeno del doblaje que obligan a “cuadrar” en el tiempo de la locución del texto con la articulación del habla, sobre todo cuando la voz es *in* y además, se percibe el rostro del actor que habla, como es la mayoría de las veces.

Finalmente el trabajo de Soto identifica la relación entre las 12 variables organizadas en adjetivos bipolares cuya interacción define la credibilidad de la voz. De sus conclusiones se desprende que:

VARIABLES como agradabilidad, credibilidad, responsabilidad, sinceridad y versomilitud, que son afirmadas para los perceptores de ambas modalidades perceptivas, descansan sobre la voz de los hablantes, y las imágenes de estos hablantes no modifican estadísticamente esta tendencia.

Los perceptores de ambas modalidades atribuyen valores imprecisos en la atribución de familiaridad, naturalidad, profundidad y simpatía. La modalidad audiovisual no modifica esta tendencia.

Sin embargo, la percepción audiovisual sí actúa sobre las tendencias generales de atribución de inteligencia y seguridad.

Habida cuenta de que algunos de los adjetivos bipolares que definen estas variables son empleados en nuestra experimentación para delimitar el carácter o actitud de los personajes analizados, y que, además, nos planteamos la comparación entre la percepción unimodal y bimodal del texto portador, las conclusiones apuntadas por Soto (2000) constituyen una valiosa referencia.

2.10.5.1 El debate sobre la percepción bimodal

La tesis doctoral de María Teresa Soto (2000) tiene además la virtud de introducirnos en el debate sobre la integración de la percepción auditiva visual, rescatando para el terreno de la Comunicación Audiovisual el legado del trabajo experimental de multitud de investigadores de la Psicología de la Percepción. Destacaremos aquí aquellas investigaciones que consideramos que arrojan luz sobre nuestro objeto de estudio y a las que hemos tenido acceso gracias a la prolija investigación bibliográfica del trabajo de Soto.

El estudio experimental sobre la integración de los estímulos visuales y sonoros viene desarrollándose desde hace décadas, principalmente en el campo de la Psicología de la Percepción. La creencia apuntada por Francis Galton hace más de cien años (citado por Harrison 1983:259) sobre la existencia de dos tipos de personas: las visualizadoras y las verbalizadores no ha sido cabalmente demostrada de manera experimental. Los estudios sobre fisiología distinguen funciones diferentes para los dos hemisferios cerebrales, el izquierdo para el proceso verbal y el derecho para funciones no verbales y aunque parece demostrado que los seres humanos difieren individualmente en la habilidad para la recepción de los respectivos estímulos, la experimentación sobre la forma en que se produce la integración auditivo visual sigue vigente.

Uno de los descubrimientos más destacados sobre la integración auditivo visual fue reportado por MacDonald y McGurk en 1978. Sus experimentos vinieron a desmentir la creencia derivada de la experimentación precedente, (fundamentalmente aquella destinada a identificar el papel de la visión como modo alternativo de oír entre sujetos con minusvalía auditiva o en condiciones de ruido) que consideraba que la comunicación cara a cara entre sujetos normales era un proceso esencialmente sonoro y que el rol de la visión era un añadido. Hasta entonces, las investigaciones experimentales sobre el papel de la visión en la percepción del habla habían arrojado hallazgos tales como proveedora de información alternativa (estudios centrados en la comparación de las habilidades para leer los labios de sujetos sordos con sujetos sin disfunción, como los de Conrad (1977) y Pelson y Prather

(1974) (Citados por MacDonald y McGurk 1978) Y, por otra parte, estudios que afirmaban el papel compensatorio o complementario de la visión en situaciones de ruido (Dodd 1977).

Las hipótesis de sus experimentos se basaban en la dificultad de discriminar visualmente las consonantes pronunciadas dentro de cada categoría: labiales o no labiales, dado que mientras el sonido es detectado por el oído, el lugar de la articulación es detectado por el ojo. Su hipótesis defendía así que la información de las dos fuentes se combina y se sintetiza.

Lo que desde entonces se conoce como efecto McGurk consiste en la evidencia demostrada por los autores citados de que ante la percepción auditiva de determinadas sílabas emitidas dobladas en sincronía con rostros que pronuncian otras sílabas, la percepción se modifica de manera que el resultado es uno nuevo.

Los investigadores experimentaron con el sonido de las sílabas “ba-ba” (consonante labial) dobladas en sincronía sobre rostros que pronunciaban “ga-ga” (consonante no labial). Entre el 80% (niños pequeños) al 98% (adultos) de los sujetos experimentales dijeron haber escuchado “da-da”. En el proceso inverso (voz ga-ga sobre articulación ba-ba) el resultado era modificado hacia la percepción “gab-ga” o “bag-ga”. En ambos casos nos encontramos con una percepción nueva.

El efecto McGurk vino entonces comprobar la hipótesis de que “la información visual del movimiento de los labios del hablante modifica profundamente la percepción del receptor del habla normal para sujetos con audición normal.” (MacDonald y McGurk 1978:253).

Este efecto se produce de manera más pronunciada cuando la combinación es de un patrón auditivo con articulación bilabial combinado con información visual de una articulación no labial.

La importancia de esta investigación reside en la demostración experimental de la integración y sintetización perceptual visual-auditiva del habla.

Para los fines de nuestro trabajo, entendemos que las condiciones de la experimentación que dieron lugar a la identificación del efecto McGurk no se reproducen estrictamente en el fenómeno del doblaje. Es decir, los receptores de una película doblada al castellano desde el inglés, no decodifican, por ejemplo, algo distinto entre el “I love you” articulado por el hablante y el “te amo” emitido por la voz sincronizada del doblador.

Ateniéndonos a la evidencia empírica descrita, es posible que la percepción estricta sí provoque el efecto McGurk, es decir, que el receptor perciba esa tercera frase entre “I love you” y “te amo”. Creemos que esta afirmación sólo puede ser comprobada mediante la experimentación que permitiera comparar la percepción del doblaje con audiencias cautivas que desconocieran tanto la lengua de origen como la de destino del doblaje percibido. Porque entendemos que factores como la escucha activa que comprende el reconocimiento de la lengua propia; lo que Chion (1990b) llama escucha causal y la semántica; aunados al aprendizaje y asunción de códigos cinematográficos, incluido el del doblaje; la percepción de sus discursos completos (*full speech*), coherentes y contextualizados, entre otros, influyen en la decodificación correcta del doblaje. De manera experimental parece demostrado que, tal como, a partir del trabajo de Massaro (1987), concluye Soto (2000:29): “debido a que el perceptor utiliza las fuentes auditiva y visual para distintos propósitos, el proceso perceptivo se salda con el realce de uno de los dos canales, más que con una solución intermedia o de compromiso. El producto perceptivo del habla bimodal es, por tanto, una sinergia del procesamiento de las informaciones sonora y visual”

Al respecto detectamos una evidencia, aunque no experimental, en nuestra investigación precedente y exploratoria sobre el fenómeno del doblaje. Entonces uno de los sujetos entrevistados relató su obvia incapacidad para decodificar películas dobladas cuyas lenguas, tanto original como de destino, desconocía. Lo relevante es que la experiencia de asistir como espectador a películas dobladas del inglés en Francia y Portugal le inducía a pensar que el doblaje en esos países era muy malo. Textualmente decía “Se les nota el doblaje, aunque cuando ves películas dobladas al español, no te estás fijando en los labios, sino en lo que te dicen” (Palencia 1998:112). La última parte de su reflexión sobre la influencia de la lengua decodificable por el conocimiento previo del receptor, y su percepción

por el canal auditivo que presumiblemente es realizado por el perceptor de doblaje, viene a avalar la hipótesis apuntada arriba.

A partir de la identificación del efecto McGurk, posteriores experimentaciones demostraron la consistencia de la integración de la percepción auditivo-visual, como acertadamente resume Soto (2000):

“Investigaciones posteriores corroboran que la integración sonoro-visual del habla bimodal es un fenómeno sólido e inmune a una variedad de situaciones experimentales. Ocurre de manera semejante entre **individuos de diferentes lenguas** (Massaro et al.: 1995, Sams et al.: 1998); **aunque el individuo perciba su lengua materna o una segunda lengua** (Reisberg et al.: 1987); **aunque la información acústica y visual se presente con un tiempo de retardo severo respecto de la información visual** (Massaro et al.: 1995); **aunque la señal acústica se presente manipulada o intacta** (Reisberg et al.: 1987); aunque la información acústica y visual no corresponda a hablantes del mismo género (Green et al.: 1991; Strand y Johnson: 1996), o a una misma fuente biológica (Green et al.: 1991); **aunque las informaciones sonoras y visuales provengan de fuentes naturales o animadas** (Massaro y Cohen: 1983); **aunque existan discrepancias entre la ubicación espacial de la señal sonora y la visual** (Green et al.: 1991); aunque se invierta el aparato de televisión que emite el rostro del hablante (Green: 1994) o aunque sea el rostro del hablante el que se presente con una rotación de hasta 180 grados sobre el eje natural de visión (Massaro y Cohen: 1996); **aunque el estímulo auditivo sean sílabas, palabras aisladas u oraciones** (Sams et al.: 1998) y **en el caso de oraciones la contribución visual es superior** (Reisberg et al.: 1987); **aunque el perceptor tenga previo conocimiento de la ilusión y considerable práctica en su detección** (Walker et al.: 1995); aunque los perceptores tengan sólo cuatro meses de edad (Burnham y Dodd: 1996) y **aunque el perceptor detecte que los estímulos no son sincrónicos** (Green: 1994). Todas estas pruebas son, en sí mismas, un ejemplo de que la presentación del efecto

McGurk avivó el interés de un grupo de investigadores interesados en el procesamiento perceptivo del habla audiovisual” (Soto 2000: 26 y 27)

Hemos subrayado en negritas las circunstancias descritas que pueden presentarse en el fenómeno del doblaje. De entre esta extensa línea de investigación nos interesa destacar los trabajos de D. Massaro y M. Cohen. En su artículo “Evaluación e integración de la información auditiva y visual en la percepción del discurso” (1983), siguiendo la línea que permitió identificar el efecto McGurk, estos investigadores prueban que la información que provee la visión del hablante, colabora fuertemente en la decodificación del discurso. Mas tarde (1987) Massaro demuestra que la percepción audiovisual es un proceso de síntesis por el cual los perceptores valoran la información que cada canal perceptivo les ofrece e inmediatamente realizan un proceso de síntesis.

Con la intención de profundizar en sus descubrimientos, en 1995 Massaro, Cohen y Smeele describen la experimentación realizada con hablantes de diferentes lenguas (alemán, inglés, japonés, español) sometidos a la percepción audiovisual de las sílabas “da” y “ba”, alternativamente. Sus resultados “desmienten el modelo de la dominancia auditiva (ADM) así como el modelo de la percepción añadida (AMP) en el cual las fuentes auditivas y visuales son combinadas linealmente”. (Massaro et al 1995:113) Por el contrario, gracias al que denominan “*fuzzy logical model of perception (FLMP)*” (modelo de percepción lógica indistinta) apoya la premisa de que “las múltiples fuentes de información continua son evaluadas e integradas en la percepción del discurso” (Massaro et al 1995:113)

Creemos que, aunque el efecto McGurk se produjera en el doblaje en los términos de su experimentación primera, es la integración y síntesis de los estímulos sonoros y visuales del habla, que todos los estudios anotados arriba prueban, lo que colabora activamente en la eficacia del doblaje. Es decir, el perceptor integraría los estímulos sonoros de las voces dobladoras con el habla visible, de manera lógica, evaluando previamente la información que cada canal le ofrece, como lo prueba el FLMP.

Así, el inevitable ruido provocado por la asincronía articulatoria del hablante en pantalla, sería desechado a favor del canal auditivo, que provee un discurso coherente y

decodificable, y subsanado por el resto de información visual, incluida, y especialmente, la interpretación gestual de los presuntos hablantes, los actores en pantalla.

Debido a la concomitancia o semejanza que muestran con el fenómeno del doblaje, nos interesa destacar las conclusiones de Massaro y Cohen (1996) referidas a la manipulación independiente de los canales auditivo y visual. Según estos autores:

“En la comunicación cara a cara, los perceptores integran el habla visible del hablante con el mensaje audible. Esta integración ha sido mostrada que ocurre incluso si las fuentes audible y visible son manipuladas de manera independiente uno del otro” (Massaro, 1987: citado por Massaro & Cohen 1996:1047)

Igualmente,

“La integración del habla visible y audible es también sorprendentemente robusta ocurriendo incluso cuando las dos fuentes no son compatibles ecológicamente. El habla audible es integrada con el habla visible sintética (animada) de la misma manera que con el habla visible natural” (Massaro, 1983 y 1990: citado por Massaro & Cohen 1996:1047) (...) “Más aún, el habla auditiva y visual es combinada de manera natural incluso cuando el sexo de la voz difiere del sexo del rostro que habla”. (Green, Kuhl, Meltzoff & Stevens 1991: citados por Massaro & Cohen 1996:1047)