



VI Congreso Iberoamericano de Acústica - FIA 2008
Buenos Aires, 5, 6 y 7 de noviembre de 2008

FIA2008-A160

Umbral absoluto de percepción de coherencia de origen entre fuente sonora y visual

Roberto Muñoz^(a),
Manuel Recuero^(b),
Diego Durán^(a),
Manuel Gazzo^(a).

(a) Ingeniería en Sonido, Universidad Tecnológica de Chile INACAP. Brown norte 290, Ñuñoa, Santiago, Chile. E-mail: rmunozs@inacap.cl

(b) Grupo de Investigación en Instrumentación y Acústica Aplicada (I2A2), Universidad Politécnica de Madrid, INSIA. Campus Sur UPM. Ctra. Valencia Km. 7. 28031 – Madrid, España. E-mail: manuel.recuero@upm.es

Abstract

This study investigated the absolute perception threshold of position coherence between auditory and visual sources. In other words, the absolute threshold, in degrees, for the maximum separation of the auditory and visual sources, which lets perceive these two stimuli as if they were coming from the same place of the screen. The experimentation was carried out in a small cinema dubbing stage. The sound source is positioned behind the centre of an acoustically transparent screen and the image is positioned on different places on the screen in steps of 1 degree. The constant stimulus psychophysical method [1] was chosen to obtain the absolute threshold, and phonetically balanced sentence was used and presented at 75 dBA. The experimentation was carried out following the standards and recommendations of the audiovisual industry; ITU775 [2] and ISO 9568 [3] and 2969 [4]. Results let develop criteria and recommendations, for the dialogs panoramic positioning for cinema audiovisual productions, based on the audience audiovisual perception capability.

Resumen

Este estudio tiene por objeto determinar el umbral absoluto de percepción de coherencia de origen entre la fuente sonora y visual en una proyección audiovisual. Es decir, se busca el umbral absoluto, en grados, para la separación de las fuentes (sonora y visual) para el cual se deja de percibir como si proviniesen ambos estímulos desde el mismo lugar de la pantalla. La experimentación se realizó en una sala de mezclas de sonido para cine, dejando la fuente sonora fija al centro mientras se movía la imagen hacia la derecha en saltos de 1°. Se usó una frase fonéticamente balanceada, presentada a un nivel de 75 dBA. El método psicofísico utilizado fue el de los estímulos constantes [1]. La experimentación se realizó cumpliendo las normas, estándares y recomendaciones de la industria audiovisual; ITU 775 [2] e ISO 9568 [3] y 2969 [4]. Los resultados permiten desarrollar criterios y recomendaciones para el posicionamiento panorámico de los diálogos en producciones audiovisuales para cine, basados en la capacidad de percepción audiovisual de la audiencia.

1 Introducción

Ciertamente, dentro del lenguaje audiovisual, el receptor está condicionado, mientras que el emisor tiene la tarea de preocuparse por cómo hacer que el mensaje pueda ser entendido. El ciclo perceptivo del lenguaje audiovisual inicia su proceso bajo las siguientes características biológicas: el estímulo es percibido mediante los mecanismos fisiológicos de la visión y la audición, los que a su vez determinan la interpretación de las sensaciones de las diversas variaciones acústicas y lumínicas del medio al cual se está sometido, continuando por un reconocimiento condicionado que se almacena de acuerdo con las características biológicas y culturales -memoria-, donde finalmente hay una respuesta del receptor. Cabe destacar que no sólo es un estímulo, sino que es un conjunto sistemático que el sujeto organiza de acuerdo a su situación contextual, donde todo receptor es capaz de sentir y percibir, como definición y características esenciales del lenguaje audiovisual.

De ahí que la correlación entre el estímulo sonoro y visual es conocida por ser crítica en la generación de un realce en la experiencia audiovisual de los espectadores, en lo que respecta a la percepción de verosimilitud de lo que se presenta.

Las evidencias que hay sobre la interdependencia entre los sentidos de la audición y la visión muestran que la sinergia perceptual depende del grado de coincidencia entre la información visual y auditiva mostrada al espectador.

La percepción auditiva y visual, no son procesos que operen aisladamente, ambas modalidades cooperan en la mejora de la habilidad y eficiencia humana de percibir el ambiente que le rodea. Cuando la información auditiva es apoyada por una coherente información visual, o cuando la información visual es reforzada por una coherente referencia auditiva, la interacción cooperativa entre ambas modalidades refuerza la comprensión del estímulo [5]

En este estudio se usó información visual correlacionada para corregir la ubicación de un sonido (efecto ventriloquismo), donde la magnitud de este efecto tiende a decrecer cuando la separación, tanto espacial como temporal, entre el estímulo visual y el auditivo, es incrementada. Este efecto puede ser aprovechado en el proceso de mezcla, con el objeto de reducir la cantidad de ajustes de posicionamiento panorámico de la fuente sonora, cuando no sea necesario, tratando de seguir a la fuente visual (los personajes se mueven constantemente a lo ancho de la pantalla).

De esta forma, la cuestión que se plantea es encontrar la distancia máxima entre fuente sonora y visual que no produzca molestia o interfiera con el concepto de “incredulidad suspendida” (Suspension of Disbelief) [6], concepto que puede ser considerado como una regla base de la producción audiovisual.

Por otra parte, el creciente aumento en el número de producciones audiovisuales, tanto para cine, vídeo y televisión ha hecho que se deban optimizar los tiempos que se requieren en las distintas etapas de producción de los mismos. Para esto se debe tener claro en el momento de la toma de decisiones, de qué manera y cuáles de los ajustes de variables y parámetros realizados por el ingeniero en sonido serán percibidos por la audiencia, en lo que respecta a la mezcla de sonido.

Cabe mencionar que en la actualidad las producciones audiovisuales se distribuyen en sistemas bien estandarizados, tanto a nivel masivo como particular. Es así como las configuraciones electroacústicas, tanto en cines comerciales como en sistemas de casa, no difieren significativamente entre un lugar y otro de reproducción, si se siguen los estándares recomendados. Esto permite que las decisiones adoptadas en la producción de sonido sean reproducidas con un alto nivel de equivalencia en el lugar de exhibición con respecto al lugar donde estas fueron realizadas. En este ámbito, últimas investigaciones avalan a la

recomendación ITU (BS.775-1) para arreglos de altavoces en sistemas de cinco canales [7]. Sin embargo, ya desde los ochenta existen claros estándares en lo que a salas de cine se refiere, tanto en los aspectos acústicos como en los electroacústicos [4]. Estos estándares deben ser seguidos por los estudios en los cuales se realizan las mezclas de sonido. En ese sentido, tanto Dolby [8], SDDS [9], como DTS [10], han sido empresas que han cuidado que se cumplan los requisitos mínimos en esta industria, donde a niveles de excelencia se ha instalado la certificación THX [11].

Hoy en día las posibilidades ofrecidas por los sistemas de mezcla de sonido para películas son muy extensas en relación al posicionamiento espacial de la fuente sonora. Las limitaciones existentes que se deben considerar están relacionadas con estándares técnicos y criterios estéticos, los que dirán cuando o no será recomendable mover la fuente sonora en a diferentes posiciones.

En este sentido es bien conocido que los diálogos en las producciones audiovisuales provienen principalmente desde el frente, esto es debido a que los personajes aparecen allí la mayoría del tiempo. Pero no necesariamente en el centro de la pantalla, ya que se están moviendo permanentemente a lo ancho de ella. ¿Será necesario seguir con el sonido cada movimiento que el personaje realiza? Sabiendo cual es la máxima separación entre fuente visual y sonora que no produce una sensación de incoherencia se podrá tener claridad a este respecto.

Se debe considerar que los lugares de escucha están pensados para un gran número de espectadores, por lo que las decisiones tomadas deben en promedio ser adecuadas para una zona amplia de escucha. Es por eso que en este estudio se realiza la experimentación en la posición más cercana a la pantalla recomendada, ya que para esa posición es para la cual se podría producir la mayor diferencia posible entre sonido e imagen. Además, debido a que existen los planos cinematográficos, en donde los personajes no aparecen siempre del mismo tamaño, se consideró recomendable la realización de dos pruebas, una con el personaje en plano general (figura 1) y otra en primer plano (figura 2).



Figura 1. Plano general



Figura 2. Primer plano

2 Metodología

2.1 Sujetos

Participan seis voluntarios (promedio de edad de 20 años), quienes informaron no tener problemas auditivos ni visuales.

2.2 Procedimiento

En este estudio las secuencias audiovisuales fueron grabadas en formato digital. Los videos fueron creados a una resolución de 720 por 480 píxeles editados en Final Cut. El audio fue editado en un sistema Protools HD3 instalado en un computador Macintosh G4.

La distribución observada de cada participante para cada experimento fue comparada con una distribución normal usando el método Kolmogorov-Smirnov con el software SPSS. Todas las observaciones podrían razonablemente provenir de esa distribución ($p > 0,05$), por lo que todos los sujetos fueron considerados en el análisis.

El porcentaje de respuestas “coherente” fue determinado con el total de los participantes. La tabla 1 muestra los resultados para el experimento en primer plano y la tabla 2 para el experimento en plano general.

Tabla 1. Resultados experimento en primer plano

Ajuste (°)	Probabilidad (%) Incoherente	Probabilidad (%) Coherente	N° de intentos	Respuestas
2	6,2	93,8	96	90
4	5,2	94,8	96	91
6	3,1	96,9	96	93
8	12,5	87,5	96	84
10	19,8	80,2	96	77
12	10,4	89,6	96	86
14	18,7	81,3	96	78
16	29,2	70,8	96	68
18	36,5	63,5	96	61
20	47,9	52,1	96	50
22	61,5	38,5	96	37
24	72,9	27,1	96	26
26	77,1	22,9	96	22

Tabla 2. Resultados experimento en plano general

Ajuste (°)	Probabilidad (%) “Incoherente”	Probabilidad (%) “Coherente”	N° de intentos	N° Respuestas “Coherente”
2	7,3	92,7	96	89
4	4,2	95,8	96	92
6	4,2	95,8	96	92
8	11,5	88,5	96	85
10	24,0	76,0	96	73
12	12,5	87,5	96	84
14	15,6	84,4	96	81
16	21,9	78,1	96	75
18	33,3	66,7	96	64
20	33,3	66,7	96	64
22	60,4	39,6	96	38
24	71,9	28,1	96	27
26	77,1	22,9	96	22

Para estimar el Umbral Absoluto, se requirió de la interpolación de los valores de la respuesta psicométrica obtenida. Para lograr esto se usó la función matemática de la ecuación 1.

$$P_{(\alpha)} = \frac{P_{\max}}{1 + \left(\frac{P_{\max}}{P_{\min}} - 1\right) e^{-KP_{\max}\alpha}} \tag{1}$$

La ecuación 2 fue usada para ajustar la curva a los datos observados

$$\alpha = \frac{1}{-KP_{\max}} \left\{ \ln\left(\frac{P_{\max}}{P(\alpha)} - 1\right) - \ln\left(\frac{P_{\max}}{P_{\min}} - 1\right) \right\} \tag{2}$$

Usando la ecuación 1, la función psicométrica que mejor se ajusta a estos datos esta dada por $K = 0,00122$, con lo que se obtiene un $R^2 = 0,97709$ y un error total de 365,5 %. Calculando el ángulo para el 50% de respuestas “Incoherente”, se obtiene un $\alpha = 20,44^\circ$.

Los resultados medidos y su aproximación matemática para ambos experimentos se pueden observar en las figuras 4 y 5. El valor obtenido para la prueba de plano general es un $\alpha = 21,02^\circ$.

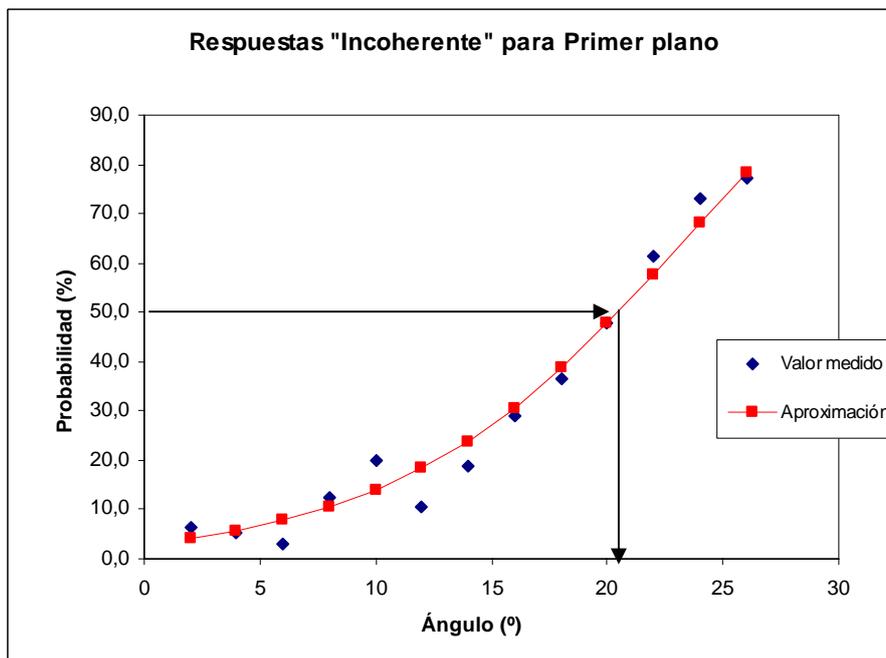


Figura 4. Gráfico de los valores medidos y su aproximación matemática para primer plano.

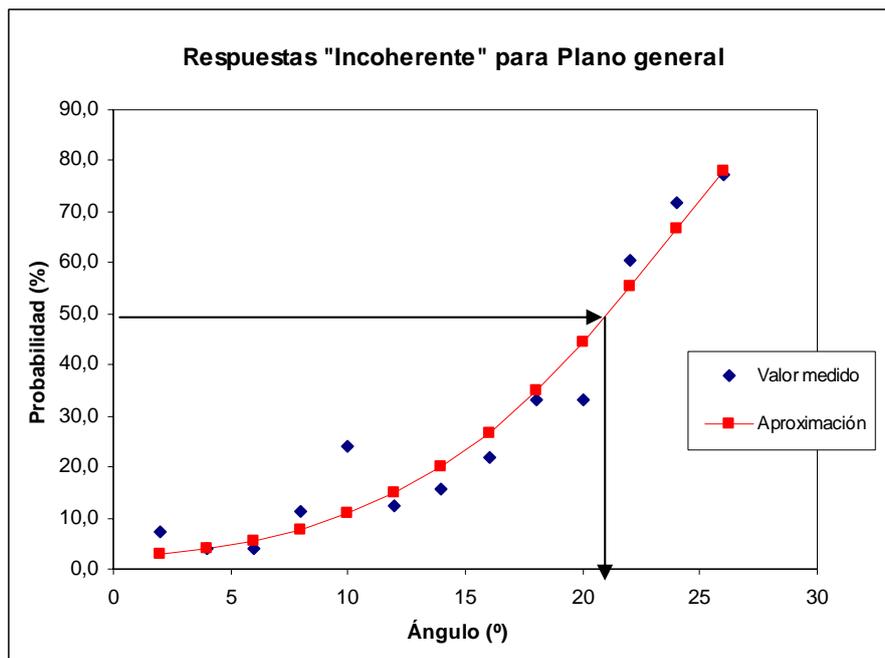


Figura 3. Gráfico de los valores medidos y su aproximación matemática para Plano general.

Un prueba t , comparando los resultados entre ambos planos, revela que la diferencia entre ellos no es estadísticamente significativa [$t(12) = 1,4$; $p = 0,093$].

4 Discusión

Como se puede observar no hay diferencia significativa entre ambos tratamientos, vale decir no influiría el tamaño de la fuente visual en la percepción de coherencia de origen entre ésta y el sonido.

Se puede concluir que no sería necesario en la mayoría de los casos realizar ajustes de posicionamiento de los diálogos aunque la imagen del personaje se mueva a lo ancho de la pantalla. Desde la posición de mezcla, el borde de la pantalla está alrededor de $22,5^\circ$ desde el centro. De esta forma el ingeniero de mezcla no percibiría incoherencia. Solo se vería progresivamente afectada, a medida que se acerca a la pantalla, la audiencia que se encuentra en el segundo tercio de la sala. Para el caso de la primera fila, el ángulo de visión desde el centro hacia uno de los costados es de alrededor de 30° , por lo que las personas sentadas en la primera fila en el eje de la sala podrían sentir incomodidad cuando la imagen está sobre los $2/3$ desde el centro hacia cualquiera de los dos costados del apantalla. Es por esto que la recomendación sería desplazar la voz desde el centro, en la misma cantidad en que la imagen sobrepasa los $2/3$ desde el centro de la pantalla.

Referencias

- [1] Gelfand, Stanley (1990). "Hearing: an introduction to psychological and physiological acoustics". Marcel Dekker Inc, New York, USA.
- [2] ITU-R BS. 775 (1994). "Multichannel stereophonic sound system with and without accompanying picture".
- [3] ISO 9568 (1993). "Cinematography -- Background acoustic noise levels in theatres, review rooms and dubbing rooms".
- [4] ISO 2969 (1987). "Cinematography -- B-chain electro-acoustic response of motion-picture control rooms and indoor theatres -- Specifications and measurements".
- [5] Woszczyk, W.; Bech, S.; Hansen, V. (1995). "Interactions between audio-visual factors in a Home Theater System: Definition of Subjective Attributes". Paper presented at the 99th AES Convention.
- [6] Taylor, Samuel (1985). "Biographia Literaria". H.J. Jackson, Oxford, UK.
- [7] Muraoka, T.; Nakazato T. (2007). "Examination of Multichannel Sound-Field. Recomposition Utilizing Frequency-Dependent Interaural cross correlation (FIACC)". Journal of the Audio Engineering Society, 55, 236-256.
- [8] www.dolby.com
- [9] www.sdds.com
- [10] www.dtsonline.com
- [11] www.thx.com
- [12] Holman, Tomlinson (2000). "5.1 Surround Sound up and Running". Focal Press, USA.
- [13] Soto, María (2000). "Influencia de la percepción visual del rostro del hablante en la credibilidad de su voz" – Tesis doctoral dpto. de comunicación audiovisual y publicidad – UAB, Barcelona, España.