

RECONSTRUCCIÓN DE ATMÓSFERAS SONORAS TRIDIMENSIONALES.

RECONSTRUCTION OF THREE-DIMENSIONAL SOUND ATMOSPHERES



Jhosimar Aguacía Fisco¹, Diego Pardo Sabogal²

*Facultad de Ingeniería de Sonido, Universidad de San Buenaventura
Bogotá, Colombia*

¹jhosimar_aguacia@hotmail.com

²died_delay@hotmail.com

Recibido: 15 /08/2013 • Aprobado: 02/09/2013

RESUMEN

El propósito de este proyecto es la reconstrucción de atmósferas sonoras tridimensionales, por medio de la tecnología Ambisonics, implementando un sistema de captura Native B-Format, codificación de fuentes puntuales y el uso de algunos audios grabados con micrófonos Soundfield, para crear así atmósferas sonoras naturales y artificiales, que serán reproducidas en un sistema 10.2, con el fin de generar un involucimiento total, incluyendo alturas.

Palabras clave: *ambisonics, atmósferas tridimensionales, B-format, codificación, decodificación, sonido envolvente.*

ABSTRACT

The purpose of this project is the reconstruction of three-dimensional sound atmospheres, through Ambisonics technology, implementing capture system Native B-Format, point source coding and the use of some audio recorded with Soundfield microphones, creating atmospheres and natural sound artificial, which will be reproduced in a system 10.2, in order to generate a total involvement, including heights

Keywords: *ambisonics, atmospheres 3D, B-format, encoding, decoding, surround.*



I. INTRODUCCIÓN

En la implementación y desarrollo de la tecnología *Ambisonics* para grabar y reproducir espacios tridimensionales se han hecho avances significativos, que permiten a esta tecnología ser una opción importante en cuanto a la grabación y reproducción de sonido envolvente, empleando métodos matemáticos matriciales. En el desarrollo de la captura, un referente muy importante y relevante para este proyecto, es lo hecho por la empresa *Nimbus Records*, la cual logra un arreglo de micrófonos que permite capturar el audio directamente en *B-Format*, siendo este, el método de captura implementado en este proyecto, eliminando, así, el proceso de codificación de la señal capturada y decodificando los ambientes sonoros directamente dentro de un arreglo de parlantes 10.2, generando la atmósfera sonora, con sonidos presentes alrededor del oyente, incluyendo alturas.

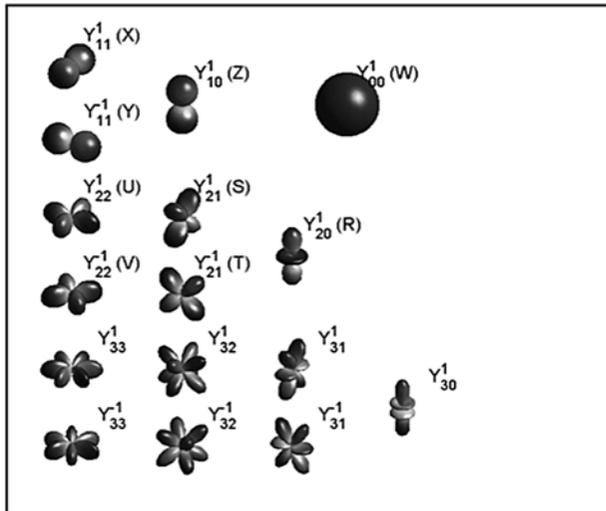
Para enriquecer el sonido obtenido por el arreglo de micrófonos se utilizan fuentes puntuales, que podrán ser codificadas y decodificadas dentro del ambiente sonoro, y ubicarse en cualquier parte al interior de la esfera sonora construida. Esta codificación de fuentes puntuales se lleva a cabo por medio del método de David Malham quien en la actualidad es uno de los principales exponentes de esta tecnología, llevando los hallazgos de Michael Gerzon, creador de la misma, a mantenerse y adaptarse a los avances tecnológicos que se presentan el día de hoy. El uso de esta codificación de señales es posible gracias a los módulos desarrollados por el ICST de la universidad de artes de Zúrich los cuales se implementan dentro del software MAX/MSP versión 6. Este grupo de módulos permite codificar y decodificar las señales en cualquier arreglo de parlantes, siempre y cuando se cumpla con los requerimientos mínimos de *Ambisonics*.

Resumiendo. Con el proceso que se lleva a cabo en la reconstrucción de las atmósferas sonoras, por medio de la tecnología *Ambisonics*, inicialmente se procede a la captura del *B-Format* y después se pasa por la codificación de fuentes puntuales y la decodificación de cualquier tipo de señal reconocida como un formato *Ambisonics* (fuentes puntuales codificadas, librerías *Ambisonics* y captura con el arreglo de micrófonos *Native B-Format*). Este proceso se realiza en el software MAX/MSP, permitiendo la reproducción de la atmósfera dentro del sistema de parlantes 10.2.

II. AMBISONICS

El sistema *Ambisonics* inicia con el matemático Michael Gerzon, quien pretendía generar sonido envolvente a partir de un arreglo de parlantes, distribuidos alrededor del oyente. Este primer sistema recibe el nombre de cuadrafónico, en el cual se disponen cuatro parlantes de forma rectangular alrededor del oyente. La captura que desarrolló este matemático contó con la ayuda de la empresa Calrec Audio, la cual creó un micrófono llamado *Soundfield* [1], micrófono que cuenta con cuatro cápsulas cardioides coincidentes, que obtienen la señal en algo conocido como el *A-Format*; sin embargo, este micrófono cuenta con una etapa de codificación que permite obtener un *B-Format* en la salida analógica del mismo.

La teoría matemática de las señales usadas por *Ambisonics*, es conocida como la teoría de los armónicos esféricos. Esta teoría de armónicos esféricos es una analogía de las series de Fourier, en las que se dice que cualquier señal puede ser reconstruida a través de la suma de muchas señales seno o coseno. Para el caso de los armónicos esféricos, este concepto sigue el mismo principio, de poder reconstruir una señal por medio de la suma de componentes individuales. (Fig. 1).



Fuente: Moreau, S., Daniel, J., & Nicol, R. (2003). Further Investigations Of High Order Ambisonics And Wavefield Synthesis For Holophonic Sound Imaging. AES 114th Convention Amsterdam.

Fig. 1 Armónicos esféricos utilizados para reconstruir las señales mediante la tecnología Ambisonics.

La cantidad de armónicos es lo que se conoce como orden en la teoría de *Ambisonics*, entendiéndose como orden cero W la función omnidireccional. El primer orden incluye el orden cero y las señales X , Y , Z , el segundo; además de estos cuatro, incluye otros cinco armónicos R , S , T , U y V , y así sucesivamente. Entre mayor sea el orden mayor será el número de armónicos. Dentro de los avances más notorios se encuentran las codificaciones fuentes puntuales, dando muy buenos resultados en la localización de la fuente, pues permite ubicar la fuente grabada en cualquier lugar alrededor del oyente de manera dinámica.

Otro aspecto importante que se debe mencionar en este artículo es el desarrollo del *Native B-Format*, por parte de la empresa Nimbus (el arreglo Nimbus-Halliday desarrollado por Jonathan Halliday y la empresa Nimbus record). Esta es una simulación del *B-Format*, formado por un arreglo de micrófonos distribuidos de la siguiente manera: un micrófono omnidireccional, preferiblemente condensador y de diafragma pequeño para el canal W , un micrófono bidireccional (Fig. 2) para X el cual apunta al frente del arreglo y un micrófono bidireccional para Y el cual debe apuntar hacia la izquierda. Para obtener la señal de altura Z es necesario incorporar un tercer

micrófono bidireccional dentro del arreglo con su cápsula apuntando hacia arriba. Aunque el arreglo original Nimbus-Halliday, solamente incorpora el eje X y el eje Y este micrófono extra está contemplado en la teoría desarrollada para este tipo de arreglos [2], [3].



Fig. 2 Arreglo de micrófonos *Native B-Format*, empleado en la captura de atmósferas sonoras, en la Universidad de San Buenaventura Sede Bogotá. (X - Y Micrófonos AKG C414, W - Z Micrófonos Marshall MXL 2010). Captura de Audio.

Este proyecto se puede dividir en dos secciones importantes, con el fin de exponer el proceso que se lleva a cabo. La primera sección hace referencia a la captura de audio. En esta parte del documento se ejemplifica la forma de obtener las señales de audio que se reproducirán posteriormente en la reconstrucción de la atmósfera sonora. La segunda sección presenta el proceso de codificación y decodificación de las señales obtenidas y la forma como estas son tratadas en el software MAX/MSP para distribuir la señal en un sistema 10.2.

A. Grabación de Audio por medio del *Native B-Format*.

La grabación de audio de este proyecto se lleva a cabo por medio del arreglo de micrófonos *Native B-Format* (Fig. 2), lo que permite capturar el ambiente y las fuentes puntuales ubicadas alrededor del arreglo.

Los ambientes que se reconstruyen son cuatro: los dos primeros constan de un juego de

baloncesto en el que se involucran tres personas; para este par de ambientes el arreglo de micrófonos está ubicado en el punto de penal y la parte inferior del tablero, respectivamente; el tercer ambiente se toma en el coliseo de la Universidad de San Buenaventura, donde se presencia un juego de voleibol. Para este tercer ambiente el arreglo de micrófonos se localiza a un costado de la red. Con estos tres ambientes se puede obtener información sonora relacionada con un plano horizontal y un plano vertical, pues se presentan rebotes en el piso y en el tablero de la cancha de baloncesto; los jugadores se mueven en el plano horizontal y adicionalmente se captura el sonido de cada entorno. El cuarto ambiente sonoro, es un ambiente artificial, que usa *B-Formats* y fuentes puntuales para simular una jungla; dentro de las fuentes puntuales más representativas se encuentran, simios y aves, que presentan un movimiento continuo a lo largo del espacio tridimensional.

El arreglo de micrófonos implementado en este proyecto consta de dos micrófonos AKG C-414 que permiten reconstruir la información relacionada con el plano horizontal X-Y y dos micrófonos Marshall MXL 2010, utilizados para capturar el componente W asociado con un micrófono omnidireccional y el componente Z, con el cual se logra la altura. Este arreglo deja grabar cualquier fuente sonora en un *B-format* (primer orden), permitiendo que las señales obtenidas se lleven al decodificador del ICST, pero antes de eso se realiza una ecualización que genera una respuesta similar a la que se obtuvo en las referencias [2] y [3].

B. Grabación de Fuentes Puntuales.

La segunda captura de audio se lleva por medio de la grabación de fuentes puntuales, la cual tiene como fin complementar el ambiente sonoro grabado y enriquecer el posicionamiento de las fuentes puntuales en el espacio, debido a que estas pueden ser codificadas hasta un tercer orden, facilitando su localización en el espacio tridimensional. Fig. 3.



Fig. 3 Grabación de las fuentes puntuales, llevadas a cabo en los estudios de grabación de la universidad San Buenaventura Bogotá.

C. Implementación de Audios B-Format, obtenidos mediante micrófonos SoundField.

Finalmente, se implementan librerías *Ambisonics*, grabadas por medio del micrófono *soundfield* y dispuestas para su uso en la página web www.ambisonia.org. Estos audios permiten incluir algunos sonidos que pueden enriquecer la atmósfera sonora; sin embargo, a diferencia de las fuentes puntuales, es posible obtener la información de todo un espacio tridimensional y, de esta manera, realizar una superposición de las atmósferas sonoras, generando un espacio dentro de otro.

III. IMPLEMENTACIÓN DEL SOFTWARE MAX/MSP PARA LA CODIFICACIÓN Y DECODIFICACIÓN DE LA SEÑAL.

En esta segunda sección se expone la manera como se implementa el grupo de módulos desarrollados por el ICST. Para comprender mejor este grupo de módulos se pueden revisar las referencias [4], [5]. Allí se describe de manera más general el uso de estos módulos y su teoría; además, se puede utilizar la información dispuesta en los ejemplos de MAX/MSP, que se incluyen en el archivo descargado de la página web del ICST. <http://www.icst.net/research/downloads/ambisonics-externals-for-maxmsp/>.

El patch que se implementa en este proyecto (Fig. 4) cuenta con dos entradas de audio destinadas a leer *B-Format*. Una de ellas, el grabado con el arreglo de micrófonos *Nat.ve B-Format* y la otra con el micrófono *Soundfield*. Estas entradas de audio se conectan directamente al decodificador *Ambisonics*, que recibe la información correspondiente a la disposición del arreglo de parlantes del sistema 10.2, por medio del ambimonitor (Fig. 5), el cual se encarga de enviar la información correspondiente a la ubicación de los parlantes o de las fuentes puntuales mediante coordenadas esféricas.

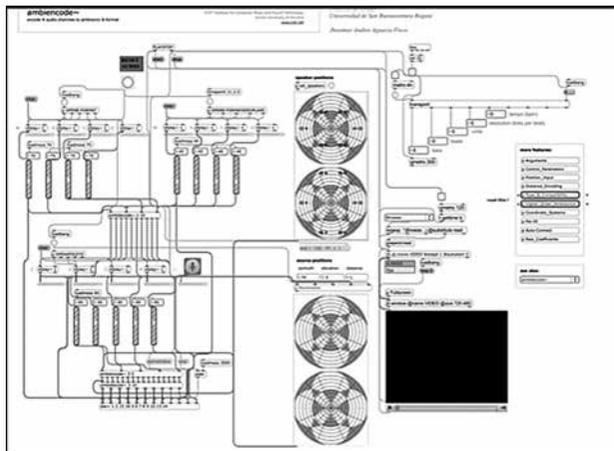


Fig. 4 Patch utilizado para reproducir las atmósferas sonoras, haciendo uso de la librería del ICST desarrollada para codificar y decodificar señales utilizando la teoría de Ambisonics, adicionando un módulo de video para sincronizarlo con el audio.

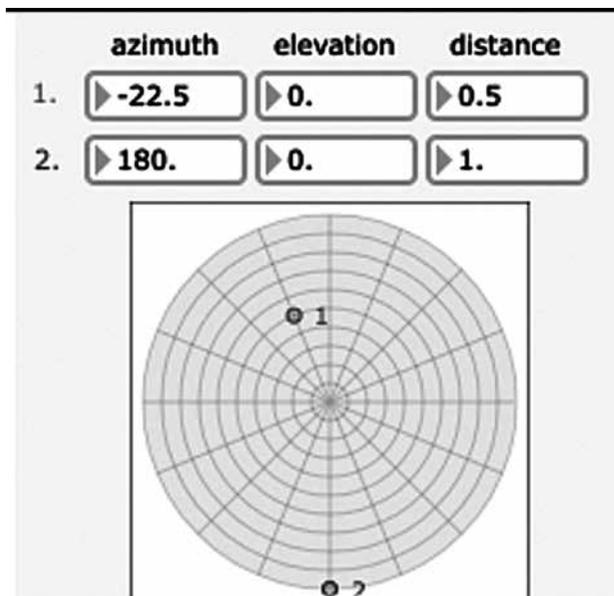


Fig. 5 Ambimonitor, utilizado en la interfaz gráfica de la librería Ambisonic del ICST.

Igualmente, se disponen otras entradas de audio que llevan la información de las fuentes puntuales: estas señales obtenidas están dispuestas en diferentes posiciones a lo largo de la reproducción de la atmósfera sonora, de acuerdo con su ubicación en el espacio tridimensional que se reconstruye. Esta información entra al codificador obteniendo de este una salida de tercer orden en el formato *Ambisonics*, que se conecta directamente al decodificador (Fig. 6).

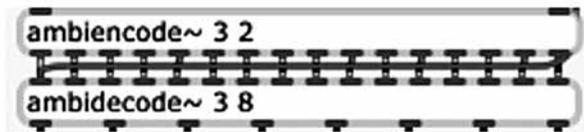


Fig. 6 Codificador y decodificador de la librería implementada por el ICST.

Por medio de estos tres tipos de señales se reconstruyen las atmósferas sonoras mencionadas en el apartado III. Así mismo, en las atmósferas sonoras relacionadas con la cancha de baloncesto y la cancha de voleibol se implementa un módulo de video que permite reproducir la atmósfera sonora de manera visual, permitiendo al oyente relacionar lo escuchado con un medio visual.

Para la reproducción de las atmósferas se utilizaron dos sistemas 5.1 idénticos, con el objetivo de garantizar una respuesta en frecuencia y un nivel de presión sonora homogéneo. Estos parlantes se ubican en una carpa hexagonal construida especialmente para el proyecto (Fig. 7). El motivo por el cual se decide usar la carpa, es para evitar reflexiones de algún espacio cerrado, intervención de la reverberación propia de la sala o problemas modales; sin embargo, estos problemas siguen estando presentes, especialmente para frecuencias altas, ya que la carpa puede presentar reflexiones, aunque de manera más tenue.

Los parlantes, siguiendo la teoría propuesta por Michael Gerzon, se colocan cada uno con un respectivo opuesto en su eje y además de esto, todos tienen una distancia de 1.5m, al centro de la carpa, (lugar donde se ubica el oyente), excepto los parlantes laterales ubicados a una distancia de 1.3m.

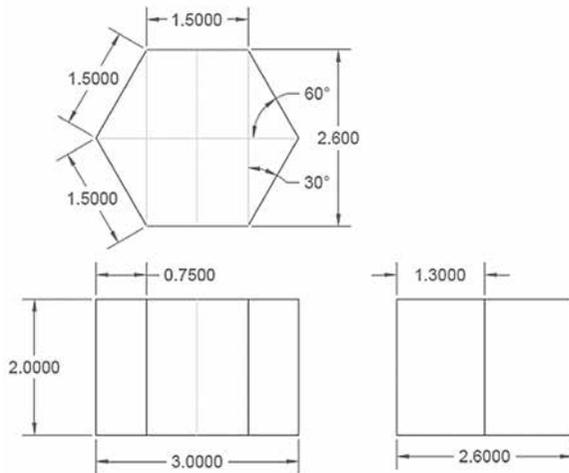


Fig. 7 (Sup.) Vistas de la carpa donde se implementa el arreglo de monitores, (Inf. Izquierda) fotografía tomada durante las pruebas y (Inf. Derecha) levantamiento 3D.

Una ventaja de trabajar con *Ambisonics* es que el arreglo de parlantes puede variar y se puede acomodar a las necesidades; simplemente se debe tener en cuenta que cada parlante tenga un opuesto y que estén ubicados a una distancia homogénea del oyente.

IV. EVALUACIÓN SUBJETIVA DEL SISTEMA

Dentro del análisis subjetivo que se lleva a cabo para evaluar este proyecto, aparte de evaluar parámetros subjetivos de un ambiente específico, se evalúa también, aunque de manera implícita, todo el proceso: la grabación, codificación, decodificación, mezcla, posproducción y disposición de los parlantes en un arreglo 10.2, ya que cada uno de estos subsistemas, que se encuentran en la cadena de procesos del proyecto, son parte fundamental del resultado final.

TABLA I.

RESULTADOS DE ANÁLISIS POR MEDIO DEL ANÁLISIS ANOVA

| Anova: Two-Factor Without Replication | | | | | | |
|---------------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| SUMMARY | Count | Sum | Average | Variance | | |
| Naturalidad | 4 | 30.27777778 | 7.569444444 | 0.120113169 | | |
| Profundidad | 4 | 30.55555556 | 7.638888889 | 0.202674897 | | |
| Presencia | 4 | 30.52777778 | 7.631944444 | 0.140110597 | | |
| Localización | 4 | 30.52777778 | 7.631944444 | 0.209940638 | | |
| Claridad | 4 | 31.94444444 | 7.986111111 | 0.034207819 | | |
| Altura | 4 | 30.5 | 7.625 | 0.285236626 | | |
| AMB 1 | 6 | 44.55555556 | 7.425925926 | 0.077988539 | | |
| AMB 2 | 6 | 46.02777778 | 7.671296296 | 0.104140947 | | |
| AMB 3 | 6 | 44.80555556 | 7.467592593 | 0.080066872 | | |
| AMB 4 | 6 | 48.94444444 | 8.157407407 | 0.019650206 | | |
| ANOVA | | | | | | |
| Source of Variation | SS | df | MS | F | P-value | F crit |
| Rows | 0.461033951 | 5 | 0.09220679 | 1.458700665 | 0.260778424 | 2.901294536 |
| Columns | 2.025977366 | 3 | 0.675325789 | 10.68357521 | 0.000519254 | 3.287382105 |
| Error | 0.948173868 | 15 | 0.063211591 | | | |
| Total | 3.435185185 | 23 | | | | |

Según la tabla I mostrada por el Anova, todos los ambientes están muy bien relacionados y, a excepción del ambiente 4, hay una similitud entre ellos. Adicionalmente, en la tabla del Anova se puede observar la similitud entre los parámetros subjetivos evaluados, teniendo todos una calificación media por encima de 7.5, lo cual es una muy buena calificación para dicho trabajo, pues representa un resumen de cualidades bien logradas para un ambiente sonoro tridimensional.

V. RESULTADOS

Por medio de esta cadena de procesos que parte de la captura de audio, pasando por un tratamiento de señales que se lleva a cabo para codificar y decodificar las señales en un formato *Ambisonics*, es posible reconstruir la señal en un arreglo de parlantes 10.2, teniendo una muy buena respuesta por parte del público.

VI. CONCLUSIONES

En la evaluación subjetiva se logró identificar y caracterizar el sistema a partir de los parámetros naturalidad, profundidad, presencia, realismo, localización, claridad, percepción de altura y

percepción de movimiento de la fuente, de cada uno de los ambientes propuestos. Viendo los resultados mediante el análisis Anova, todos los ambientes tienen una similitud entre ellos, a excepción del ambiente 4, el cual muestra un valor p o valor de probabilidad por debajo de 0.05, en comparación con los otros ambientes. Este valor significa que hay una diferencia significativa respecto de los otros 3 ambientes, en su valor medio. Además, en la tabla del Anova se puede observar la similitud entre los parámetros subjetivos evaluados, teniendo todos una calificación media por encima de 7.5, lo que representa, en resumen, cualidades bien logradas para un ambiente sonoro tridimensional.

El sistema que se implementó fue bien acogido por el público, conociendo con antelación resultados de algunos estudios hechos anteriormente con sistemas *Ambisonics*. Este sistema no es una excepción y, por el contrario, funciona correctamente. De acuerdo con lo obtenido en las encuestas, se obtuvieron calificaciones de 7 y superiores, en cada uno de los parámetros evaluados, lo cual ubica este proyecto dentro de los estándares de evaluaciones subjetivas, realizadas por otros investigadores.

El ambiente que más gustó a las personas encuestadas, fue el ambiente cuatro. La mitad de las personas votaron por dicho grupo, debido a que disfrutaron mucho la percepción auditiva que les brindaba dicho ambiente, sin tener alguna distracción visual. Además, este ambiente tiene muchas más fuentes sonoras que los otros tres y presenta múltiples movimientos en el espacio, lo que produce en el oyente una sensación más agradable, pues se tienen movimiento en el plano horizontal y vertical.

La integración de las diferentes posibilidades que brinda el sistema *Ambisonics*, como lo son

codificación y decodificación de fuentes puntuales, captura por medio del arreglo Nimbus-Halliday y el uso de librerías *Ambisonics*, le da a las atmósferas grabadas y a las creadas, variedad sonora y permite enriquecer la proyección audiovisual.

El objetivo principal de recrear atmósferas tridimensionales usando sistemas *Ambisonics* se logró, pues fue posible tener espacios donde las personas percibieron altura, profundidad, naturalidad y otros aspectos relacionados con el sonido tridimensional. Esto implica que la grabación, la codificación y la decodificación y la reproducción del sistema funcionaron correctamente, logrando integrar todos estos aspectos en un resultado final.

Aunque el sistema *Ambisonics* es una buena opción y se obtuvieron resultados positivos, esto no quiere decir que sea mejor que algún otro sistema *surround*. Simplemente se puede decir que es un sistema que funciona y puede ser implementado dependiendo de las necesidades que se quieran suplir.

REFERENCIAS

- [1] K. Farrar, *Soundfield microphone*. Wireless World, 1979, pp 48-50, 99-103.
- [2] E. Benjamin and T. Chen, "The Native B-Format Microphone: Part I," *AES 119th Convention*. New York, 2005.
- [3] E. Benjamin and T. Chen, "The Native B-Format Microphone: Part II," *AES 120th Convention*. Paris, 2006.
- [4] D.G. Malham, "Computer control of ambisonic soundfields," *AES 82nd Convention*. Londres, 1987.
- [5] M. Neukom, "Decoding Second Order Ambisonics to 5.1 Surround Systems," *AES 121st Convention*. San Francisco, CA., 2006.
- [6] Rimell, A., & Hollier, M. Reproduction of spatialised audio in Immersive Environments with Non-Ideal Acoustic Conditions. *AES 103rd Convention*. New York, 1997.

