

El Vocoder fue el primer compresor-descompresor de voz que se quiso utilizar para transmisiones telefónicas trasatlánticas. Pretendía optimizar los recursos físicos limitados, los cables, comprimiendo la voz para transmitir más comunicaciones por ellos.

Homer W. Dudley fue el técnico de los laboratorios Bell encargado de desarrollar el proyecto. El artículo describe las etapas de análisis y razonamiento que empleó Dudley. Con el estudio de la articulación vocal, el análisis espectral de los sonidos y unas imaginativas propuestas consiguió que el aparato funcionase. El Vocoder fue utilizado como sintetizador de voz y favoreció los estudios de fonética.

*Inmediatamente se sintió sorprendido ante la naturaleza de la voz del prototipo LNE. Tenía un timbre como jamás lo tuvo ninguna voz de robot que él hubiera oído (y había oído muchas). Formaba sílabas como el campanillo de un carillón de tono bajo. Esto era tan sorprendente que el técnico tardó unos momentos en oír, las sílabas formadas por esos sonidos celestiales.
Eran: «Da, da, da gu».*

«Lenny». Isaac Asimov.
(en *Los soñadores expertos* 1969)

El Vocoder: la voz de la lluvia

Luis Fernando Real Martín,
Ingeniero Técnico de Telecomunicaciones, lrealmar@gmail.com

El Vocoder fue un compresor de voz analógico para canales telefónicos construido en los años treinta. ¿Qué originales hipótesis sobre la voz humana pudieron llevarlo a cabo?, ¿cómo era el diseño electrónico que lo materializó?

En los años veinte, AT&T, ante la creciente demanda de servicios telefónicos, deseó extender la red telefónica a través del Atlántico con cables submarinos. Inventaría el primer compresor-descompresor de voz, el Vocoder para aumentar el número de conversaciones en las líneas. El aparato abrió las puertas a otras nuevas investigaciones: la síntesis artificial del habla. ¿Cómo se desarrolló este importante hito científico y tecnológico?

CABLES SUBMARINOS Y COMPRESORES DE VOZ

Las comunicaciones telefónicas de principios de siglo XX demostraron que no era necesario transmitir todo el ancho de banda vocal para entender los mensajes. Si la energía se concentra principalmente entre 200Hz y 3.000 Hz: ¿cuánto más se podría comprimir el espectro sin perder inteligibilidad?, ¿sería técnicamente posible? AT&T deseaba responder a estas preguntas y afrontar el nuevo reto para desarrollar su proyecto de redes transatlántica. En 1928 encargó a Homer W. Dudley que investigase la compresión de voz en el departamento «Speech and Hearing» de los Laboratorios Bell. A pesar de los esfuerzos, hubo que esperar a



Retrato de Homer W. Dudley. Por el Dr. John Ohala. Cortesía del Departamento de Lingüística de la Universidad de Estocolmo.



Figura 1. Compresor y descompresor de voz en una línea telefónica.

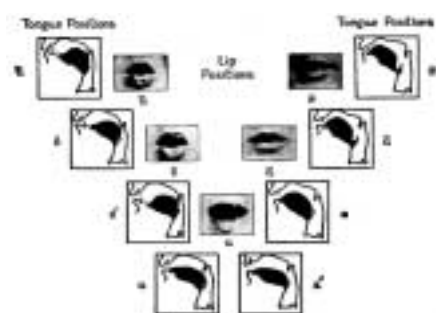
que la invención del aislante de polietileno protegiese los cables de la acción del agua marina. La red submarina fue posible a partir de 1956.

En la red telefónica, los caminos de transmisión y recepción necesitan separarse para comprimir en un extremo y descomprimir en el otro, Figura 1. Esto incrementaría la complejidad y el coste del proyecto.

MODELOS DE SISTEMA VOCAL

El primer paso de Dudley fue conocer los órganos fonadores humanos. La fonética de la época era conocida como *articulatoria*, porque se creía que la posición y el lugar donde los órganos se colocan durante la pronunciación, «punto de articulación», era la base del habla (la teoría de formantes o fonética acústica la sustituyó en los años cincuenta).

Las partes fisiológicas se dividen en dos grupos: los elementos variables o activos, cuya posición varía desde la producción de un sonido a otro como son los labios, los dientes o la lengua y los elementos fijos como la cavidad nasal, la laringe o la faringe. Las cuerdas vocales se



Posición de labios y lengua para sonidos vocálicos.
Fletcher, Harvey. *Speech and hearing*. Ed. D. Van Nostrand Co. Nueva York, 1929.

sitúan en ambos. La vibración (tono laríngeo o frecuencia fundamental) constituye un elemento fijo pero las variaciones de la tensión muscular, que modifica la frecuencia, son elementos variables.

La hipótesis de Dudley para afrontar la posibilidad de la compresión consideraba que el habla tiene un grado de predicción que genera el oyente según va escuchando. Este factor predecible se podría suprimir en el transmisor y regenerar en el receptor, antes de entregar el mensaje al oyente. La consecuencia sería la disminución de la información y el ancho de banda. El problema era identificar dicho factor. ¿Qué propiedades constitutivas del habla son prescindibles?

EL MODELO ELÉCTRICO

Harvey Fletcher describió en su libro *Speech and Hearing* publicado en 1929, un circuito eléctrico similar al órgano fonador: la señal de un oscilador armónico, elemento fijo, era moldeada por filtros que se seleccionaban con conmutadores, elementos variables, para producir el sonido adecuado.

ANÁLISIS DEL FENÓMENO VOCAL

Dudley procedió a analizar la voz, no desde su modo de articulación sino desde sus resultados; es decir, de la distribución de la energía espectral de los sonidos, gracias a su conversión de éstos en señales eléctricas, como sugería el modelo de Fletcher. Encontró dos tipos de sonidos: los «sonoros» y los «sordos». En el fenómeno de la sonoridad participan las cuerdas vocales; por el contrario, en la sorder

no, su componente principal es ruido (por ejemplo, los sonidos fricativos u oclusivos). El símil eléctrico de Fletcher se complicaba, era necesario añadir un generador de ruido para los sonidos sordos.

Dudley dividió el espectro completo, Figura 2, en dos modelos conceptuales:

- 1) El *espectro discreto* como una subbanda acotada alrededor el tono laríngeo y sus armónicos.
- 2) El *espectro continuo* como un conjunto de subbandas que divide la banda vocal. Consideró *a priori* diez subbandas.

El siguiente paso fue el análisis dinámico, la variación de la energía sonora en las subbandas a lo largo del tiempo. Como resultado enumeró una serie de características constantes y variables. Entre las primeras:

- a) Una distribución de energía sensiblemente constante en determinadas subbandas.
- b) La presencia de una frecuencia espectral que alterna desde el espectro continuo al discreto con ligeras variaciones de la posición de la fundamental.
- c) Las variaciones de la frecuencia fundamental y la energía ocurren en los intervalos silábicos.

Entre las segundas:

- d) Cambios bruscos del promedio de energía en cada subbanda.
- e) La naturaleza de la onda de tono laríngeo que es una magnitud de la frecuencia fundamental tanto en generación del espectro continuo como en el discreto.

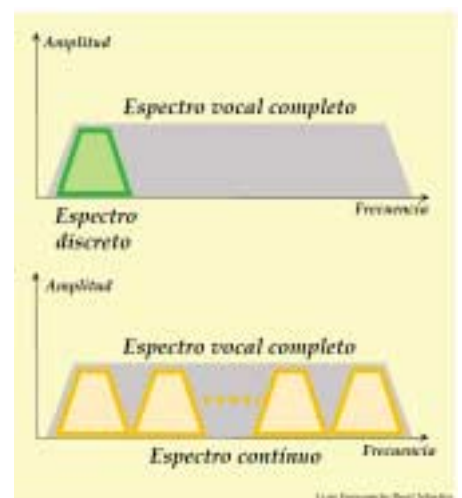


Figura 2. Modelos espectrales definidos por Dudley.

Dudley había encontrado los elementos prescindibles: eran las características constantes de la dinámica del habla que aparecían en sus modelos espectrales. Sería suficiente transmitir solo las variaciones de determinados parámetros extraídos de la voz. En el circuito compresor, un «anizador» realizaría esta función. En el descompresor, un «regenerador» aportaría las características constantes para completar el mensaje sonoro al oyente.

ESBOZANDO EL CIRCUITO

Dudley definió unos «patrones» funcionales que le ayudasen a concebir el circuito analizador: el patrón de frecuencia y el patrón de amplitud. El patrón de frecuencia distribuirá la energía sonora según el modelo espectral y el patrón de amplitud lo hará según el modelo continuo. En el receptor, la actuación conjunta de ambos patrones reconstruirá el espectro original. Concibe estos patrones como módulos que simplificarán el diseño de los circuitos (bloques verdes y naranjas de las figuras).

LA COMPRESIÓN DE LA VOZ

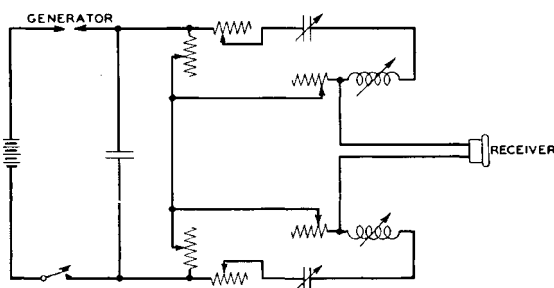
Pero... ¿Cuál es el ancho de banda mínimo en que puede comprimirse la voz manteniendo la inteligibilidad del mensaje reconstruido? La calidad del mensaje se entendía como la identificación del sujeto hablante por parte del oyente. Dudley vuelve a recurrir a los fonetistas. Primero, partiendo de los rasgos fonéticos como son los tipos de sonidos: vocales, semivocales y las consonantes; los modos de articulación: oclusivas y fricativas; o la sonoridad: sonoras y sordas, enumeró los parámetros que los crean:



Retrato de Harvey Fletcher, hacia 1975. Cortesía de Thomas H. Fletcher.

- 1) La presión del aire durante la espiración.
- 2) Las cuerdas vocales.
- 3) La parte posterior de la cámara bucal.
- 4) La parte anterior de la cámara bucal.
- 5) La apertura y cierre de atrás hacia delante de la cámara bucal.
- 6) La apertura de la boca.
- 7) La cámara de resonancia nasal.
- 8) La posición del velo del paladar y la úvula.
- 9) La posición y oclusión de los labios y dientes.
- 10) Apertura de la boca en las consonantes fricativas.

Estos parámetros los tomó como variables independientes que intervienen en la producción de los sonidos. Regulan la cantidad de energía en relación con la frecuencia y su actuación queda reflejada en las subbandas. Las diez subbandas que consideró *a priori* surgen como una «equivalencia» de estas variables.



Esquema de un circuito para producir sencillos sonidos vocálicos. Fletcher, Harvey. *Speech and hearing*. Ed. D. Van Nostrand Co. Nueva York, 1929.

El siguiente paso fue averiguar cuánto tiempo invierte cada elemento activo en completar un ciclo de cambio de posición. Fletcher, años antes, ya filmó películas del movimiento de los labios, los dientes, la lengua, etc. durante el habla. Efectuaban hasta 10 ciclos por segundo.

El producto de las variables independientes (10) por el tiempo (10 cps) proporciona las combinaciones posibles en que pueden interactuar. El ancho de banda mínimo para transmitir toda la información sería 100 cps. Las pruebas posteriores en el laboratorio demostraron valores adecuados desde 350 cps hasta 600 cps. Por lo tanto, el espectro vocal podía reducirse a 350 Hz.

PROYECTO «VOCODER»

El proyecto se denominó Vocoder, *Voice Operated reCORDER* y llegó a patentarse con el n° US 2.151.091, Figuras 3-A y 3-B. La descripción es la siguiente.

EL CIRCUITO ANALIZADOR Y COMPRESOR

La voz del micrófono, Figura 4-A, se divide en dos ramas. La PF_0 se rige como patrón de frecuencia (verde). La frecuencia fundamental se obtiene del filtro F_0 .

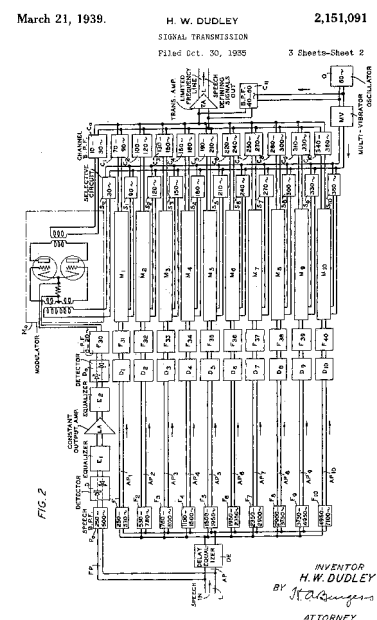


Figura 3-A. Circuito compresor. Patente n° 2.151.091

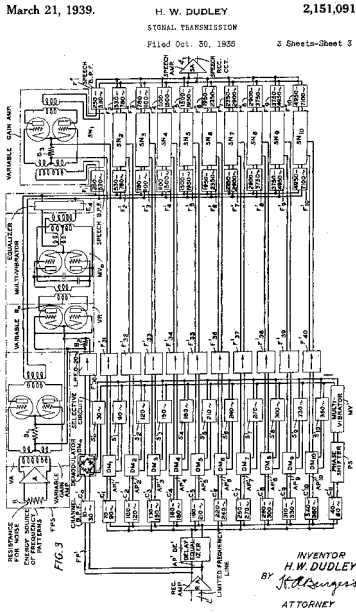


Figura 3-B. Circuito descompresor. Patente n° 2.151.091

El ecualizador E_0 y el detector D_0 realizan una conversión de frecuencia en tensión. Un filtro F_{S0} acota el ancho de banda.

En las ramas de amplitud, PA_1 - PA_{10} (naranja), el ecualizador E_1 mantiene el nivel original del micrófono. Los filtros F_1 - F_{10} crean las subbandas y los detectores D_1 - D_{10} de cada una traducen las variaciones de energía en tensión.

En el compresor (bloques azules), las tensiones de los canales modulan las portadoras, f_0 - f_{10} . Las portadoras se extraen

filtrando la salida de un multivibrador controlado por un oscilador patrón de 60 Hz. Los filtros C_0 - C_{10} evitan los solapamientos. Las portadoras moduladas se suman en el amplificador AL para transmitir este nuevo espectro reducido por el enlace telefónico. Para la demodulación se envía una señal patrón de 60 Hz.

EL CIRCUITO DESCOMPRESOR Y REGENERADOR

El receptor, Figura 4-B, divide la señal telefónica en los mismos canales. Un filtro selectivo, F_{0s} , extrae la señal patrón para estimular un multivibrador que regenere las portadoras f_0 - f_{10} . Los filtros C_0 - C_{10} separan cada canal y los demoduladores DM_0 - DM_{10} entregan las variaciones de las subbandas originales.

Las dos «fuentes patrones» de frecuencia son el multivibrador MV_0 («buzzer») y el generador de ruido térmico («hisser»), Arg . Pueden intervenir separadas o juntas. Cuando el sonido recibido es sonoro la salida en F_0 activa MV_0 a la vez que anula a Arg . Si el sonido es sordo, sin frecuencia fundamental, no hay salida en F_0 e interviene el ruido de Arg . La frecuencia patrón seleccionada se distribuye por todas las subbandas a través de los filtros F_0 - F_{10} y se suma a la señal demodulada en los amplificadores AV_1 - AV_{10} . La señal en cada subbanda dejará



El Vocoder en Alemania en los años cincuenta.

pasar una parte proporcional de frecuencia que proviene de la fuente patrón. La recuperación total del ancho de banda se realiza en el amplificador AL y se entrega a un altavoz.

«LA VOZ DE LA LLUVIA»

Hacia 1939 Dudley simplificó el Vocoder, Figuras 5 y 6. Suprimió los moduladores y demoduladores e investigó la regeneración vocal. Para la correcta reconstrucción del espectro los filtros del regenerador deben ser iguales a los del analizador, pero ¿qué ocurre si se altera esta concordancia? Hubo muchos experimentos. Cuando un canal se intercambiaba por el adyacente provocaba la confusión entre algunas sílabas. Si algunas subbandas se



Artículo sobre el «Vocoder» en la revista Science News Letter en enero de 1939. Cortesía del archivo fotográfico del Science Service de Smithsonian Institution.

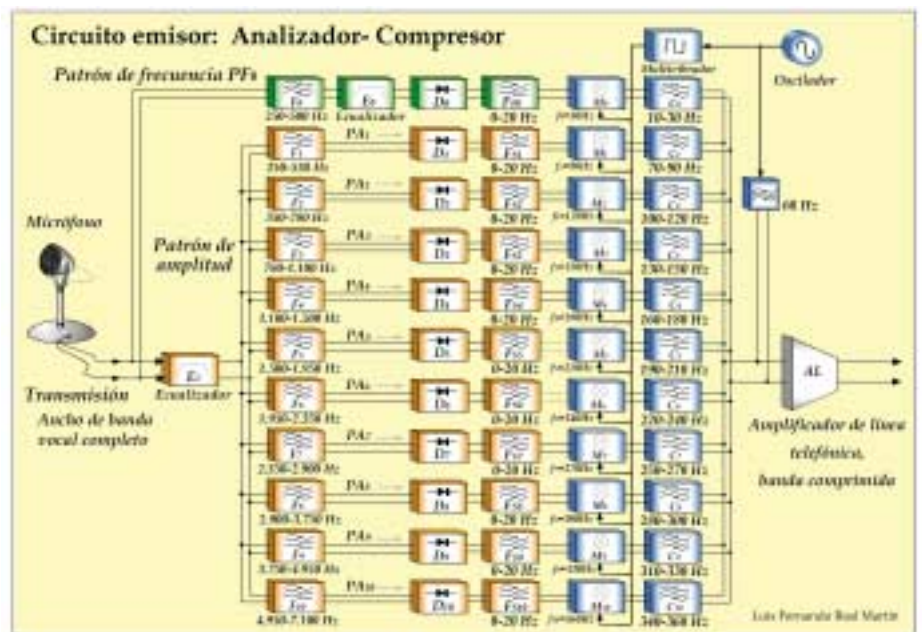


Figura 4-A. Circuito compresor simplificado.

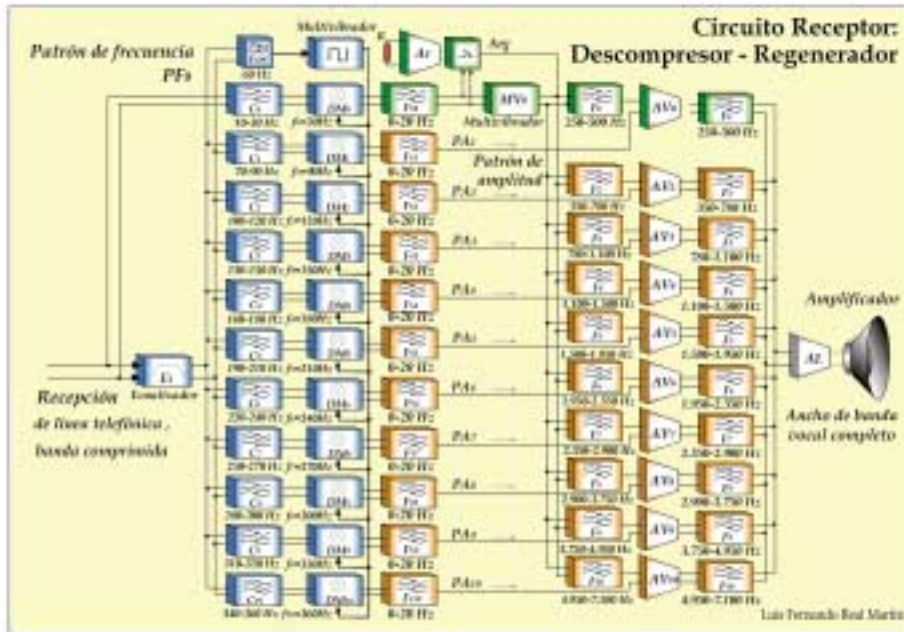


Figura 4-B. Circuito descompresor simplificado.

suprimían, la voz se escuchaba sin timbre, «artificial». Variando la frecuencia del oscilador patrón o alterando la banda de paso de los filtros, la energía aparecía en tonos más agudos o más graves, resultaba que una voz «varonil» podía convertirse en «femenina» o «infantil» y viceversa. Los filtros podían modificar la voz hasta convertirla en susurros o gruñidos.

El Vocoder permitió mezclar sonidos. Las fuentes de frecuencia se podían sustituir por un tocadiscos u otro micrófono. Los sonidos de la lluvia, del oleaje, las cataratas, el motor de un avión, el trino de aves o instrumentos musicales, previamente

grabados en un disco, eran modelados por la voz humana creando efectos que poéticamente describió Dudley como la «voz de la lluvia» ó las «conversaciones de las hojas de otoño». Propuso su aparato para la creación de efectos sonoros en el cine.

COMUNICACIONES SECRETAS Y MÚSICA ELECTRÓNICA

La función compresora fue abandonada en los años 40 por la novedosa modulación por pulsos codificados (*Pulse Code*

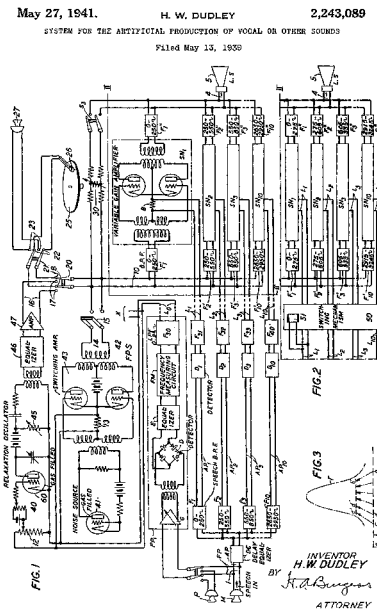


Figura 5. El Vocoder de 1939. Patente nº 2.243.089.

Modulation, PCM). Durante la Segunda Guerra Mundial, el Vocoder, junto con circuitos PCM, participó como encriptador de comunicaciones secretas del ejército aliado. Después de la guerra, el Vocoder en Europa, se unió a la pléyade de instrumentos electrónicos musicales contribuyendo a la nueva exploración sonora y estética.

A partir del diseño del Vocoder, Dudley creó el primer sintetizador verdadero de voz: el «Voder». Pero esto son otras historias.

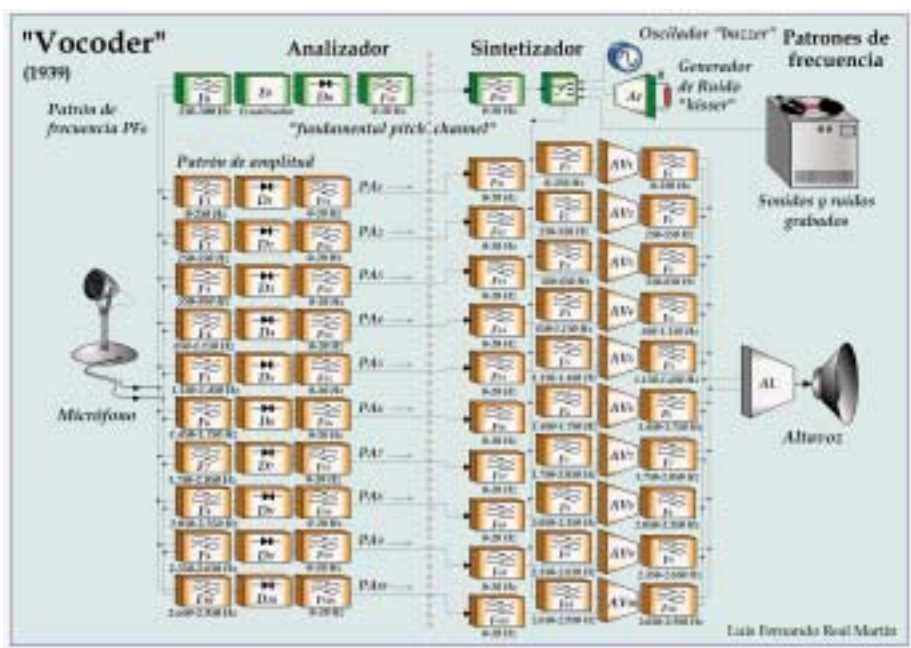


Figura 6. Diagrama simplificado del Vocoder de 1939.

REFERENCIAS

- AT&T. <http://www.att.com/history/index.html>
- BELL SYSTEM. <http://www.bellsystemmemorial.com/about.html>
- BELL TELEPHONE LABORATORIES. <http://www.bell-labs.com/about/history/index.html>
- DUDLEY, Homer W. Estados Unidos patente núm. 2.151.091. «Signal transmission». Asignada a Bell Telephone Laboratories, Nueva York, Nueva York. Solicitud: 30 octubre 1935. Pública: 21 marzo 1939.
- DUDLEY, Homer W. Estados Unidos patente núm. 2.243.089. «System for the artificial production of vocal or other sounds». Asignada a Bell Telephone Laboratories, Nueva York, Nueva York. Solicitud: 13 mayo 1939. Pública: 27 mayo 1941.
- FLETCHER, Harvey. *Speech and hearing*. Ed. D. Van Nostrand Co. Nueva York, 1929. 331 páginas.
- MARTINEZ Celdrán, E. *Fonética*. Editorial Teide, Barcelona 1994.