



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 357 700**

51 Int. Cl.:
G10L 13/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **01991746 .7**

96 Fecha de presentación : **21.11.2001**

97 Número de publicación de la solicitud: **1344211**

97 Fecha de publicación de la solicitud: **17.09.2003**

54 Título: **Dispositivo y procedimiento de edición diferenciada de voz.**

30 Prioridad: **20.12.2000 DE 100 63 503**

45 Fecha de publicación de la mención BOPI:
28.04.2011

45 Fecha de la publicación del folleto de la patente:
28.04.2011

73 Titular/es: **BAYERISCHE MOTOREN WERKE
AKTIENGESELLSCHAFT
Petuelring 130
80809 München, DE**

72 Inventor/es: **Obert, Georg y
Bengler, Klaus**

74 Agente: **Lehmann Novo, María Isabel**

ES 2 357 700 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Dispositivo y procedimiento de edición diferenciada de voz.

5 La presente invención concierne a un dispositivo de edición o generación diferenciadas de voz y a un procedimiento correspondiente, a sistemas para uso con el dispositivo de edición de voz y a combinaciones de un dispositivo de edición de voz con al menos dos sistemas, especialmente para uso en un vehículo.

10 En los vehículos se utilizan sistemas individuales que disponen de una interfaz acústica hombre-máquina para la edición de voz. En estos sistemas está siempre asociado directamente un módulo de edición de voz. Los procedimientos de producción de voz empleados se basan casi siempre en modulación de código de impulsos (= PCM), en donde se puede conectar una compresión subsiguiente (por ejemplo, MPEG). Otros sistemas emplean procedimientos de síntesis de voz que forman principalmente palabras y frases (manipulación de señales) a través del ensamble de segmentos de sílabas (fonemas).

15 En los procedimientos de edición de voz citados existe también una dependencia del hablante que requiere que, al ampliar el volumen de palabras o de texto, se soliciten tomas una y otra vez al mismo hablante humano. Además, los procedimientos PCM, exactamente igual que una síntesis de fonemas de alto valor cualitativo, requieren, debido a la manipulación de señales, un espacio de memoria considerable para archivar textos o segmentos de sílabas. En ambos procedimientos el espacio de memoria aumenta aún considerablemente cuando se deben editar diferentes idiomas nacionales.

20 Asimismo, se conocen procedimientos que se basan en una síntesis completa de la voz. Se conocen especialmente procedimientos que convierten el tracto vocal humano en el equivalente eléctrico y trabajan con un generador de sonido y varios filtros pospuestos (modelo fuente-filtro). Un aparato que trabaja según este procedimiento es un llamado sintetizador de formantes (por ejemplo, KLATTALK). Este sintetizador de formantes tiene la ventaja de que se puede influir sobre las propiedades características de la voz.

25 El documento EP-A-0 901 000 describe un dispositivo de tratamiento de mensajes con medios de recepción para recibir mensajes enviados, una memoria para almacenar una pluralidad de articulaciones diferentes (tono de voz o tono vocal) y medios de asignación para asignar una articulación de la pluralidad de articulaciones a al menos un mensaje recibido. Otra articulación es asignada a otro mensaje recibido y unos medios de edición editan el primer mensaje con una primera articulación y el segundo mensaje con una segunda articulación.

30 RUTLEDGE J C ET AL: "SYNTHESIZING STYLED SPEECH USING THE KLATT SYNTHESIZER", PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ACOUSTICS; SPEECH; AND SIGNAL PROCESSING (ICASSP): DETROIT; 9-12 DE MAYO DE 1995. SPEECH, NUEVA YORK, IEEE, US, Vol. 1, 9 de Mayo de 1995 (09-05-1995), páginas 648-651, XP000658077 ISBN: 0-7803-2432-3, se ocupan en general del funcionamiento y los conocimientos experimentales en la construcción de sintetizadores de la voz basados en el tracto vocal (sintetizador Klatt) para la síntesis de diferentes estilos de voz con diferentes características de la voz.

35 La invención se basa especialmente en el problema de proporcionar un dispositivo de edición de voz central con un gran número de sistemas en el que se active por los sistemas un único generador de voz con una pequeña memoria de parámetros.

Este problema se resuelve según la invención con las características de la reivindicación 1. Ejecuciones ventajosas son objeto de las reivindicaciones subordinadas.

40 La invención tiene la ventaja de que con un único dispositivo de edición de voz o dispositivo de síntesis de voz son posibles ediciones de voz para sistemas diferentes, pudiendo identificarse cada sistema por diferencias características de la voz.

45 Según una forma de realización preferida de la invención, se ha asociado a cada sistema un juego de parámetros que es empleado por el dispositivo de síntesis de voz para realizar una edición de voz desde este sistema. Por ejemplo, se proporciona un primer juego de parámetros para un ordenador de a bordo, un segundo juego de parámetros para un sistema de navegación, un tercer juego de parámetros para informaciones del tráfico, un cuarto juego de parámetros para un sistema TTS (sistema de texto a voz), tal como correo electrónico, y uno o más juegos de parámetros adicionales para sistemas adicionales.

50 Dependiendo del juego de parámetros asociado, el dispositivo de síntesis de voz genera la edición de voz, por ejemplo con una suave voz femenina, por ejemplo para ediciones de voz de un sistema de navegación, o con una fuerte voz de bajo masculino, por ejemplo para la edición de voz de avisos del tráfico.

55 Según una forma de realización preferida de la invención, se emplean un procedimiento y un dispositivo para una síntesis de voz completa, preferiblemente un sintetizador de formantes. En este caso, se dividen en clases los parámetros de control para el sintetizador. Una clase de parámetros dinámicos controla la articulación, tal como el movimiento del tracto vocal al hablar. Una segunda clase de parámetros estáticos controla propiedades características del hablante, tales como la frecuencia fundamental del generador y los formantes fijos que, en el caso de un niño, una mujer o un hablante masculino, son formados por la diferente dimensión geométrica del tracto vocal.

60 En un modelo ampliado del sintetizador de formantes es posible una generación separada de sonidos con voz y sin voz. En este caso, se pueden intercalar mediante otros parámetros resonadores o miembros de amortiguación adicionales o bien se puede influir sobre los parámetros dinámicos para la articulación.

El dispositivo según la invención y el procedimiento según la invención se pueden utilizar especialmente en

sistemas de un vehículo. Cada sistema tiene, por cada edición de voz, dos posibilidades de controlar la edición de voz. La primera posibilidad de edición de voz comprende la emisión de una secuencia de órdenes de control para la articulación de voz, estando almacenada en el sistema la secuencia de los parámetros de control para palabras, frases y series de frases. La segunda posibilidad de control de la edición de voz se efectúa a través de una segunda edición que conmuta un juego de parámetros que es determinante de la característica del hablante.

Como alternativa o adicionalmente, es posible también almacenar este juego de datos paramétricos directamente en el sistema y cargar el juego de datos paramétricos en el dispositivo de síntesis de voz en caso de una edición de voz necesaria.

Según otra forma de realización preferida, que se puede utilizar como alternativa o adicionalmente a las formas de realización anteriores, se pueden variar, además, dinámicamente los parámetros del generador y de los formantes para diferenciar las fuentes de información, es decir, los sistemas que realizan una edición de voz. Se pueden conseguir así diferencias audibles en la prosodia, tales como la duración y/o la acentuación de segmentos de sílabas y/o la melodía de las frases. En especial, se puede utilizar una modulación prosódica en función, por ejemplo, de una condición de tráfico o una situación de tráfico para la edición de voz de textos de anuncios. Por último, la importancia de una información puede expresarse por modulación de la voz.

La invención tiene la ventaja de que, por ejemplo, en un vehículo solamente un único generador de voz con una pequeña memoria de parámetros puede ser activado desde varias fuentes de información. Las fuentes de información pueden equiparse aquí con características de voz diferentes.

Cuando se utiliza un dispositivo de síntesis completa, por ejemplo un dispositivo de síntesis de tracto vocal, se tiene que el procedimiento es independiente del hablante y no se necesitan grabaciones de estudio de alto valor.

En un sintetizador de formantes ampliado se puede incorporar también según la invención una expresión emocional en la voz.

Con ayuda de plantillas de parámetros prefabricadas se puede variar muy fácilmente la característica de la voz. El procedimiento es adecuado también para la conversión de textos libres en voz (texto a voz), por ejemplo para la lectura del correo electrónico.

Se explica seguidamente la invención con más detalle ayudándose de un ejemplo de realización y del dibujo.

La figura 1 muestra una representación de principio de una forma de realización preferida de la invención para la edición diferenciada de voz con varios sistemas según la invención.

La forma de realización preferida de la invención representada en la figura 1 presenta una unidad de edición de voz 1 con un dispositivo de síntesis de voz 10 que en el ejemplo es un módulo de síntesis de tracto vocal y que se basa en una síntesis de voz completa. Por ejemplo, se puede utilizar un sintetizador de formantes como el KLATTALK. El dispositivo de síntesis de voz 10 está unido con un amplificador 12 cuya salida 14 suministra una señal de audio que edita voz a través de un altavoz (no representado). El dispositivo de síntesis de voz 10 lleva asociados N juegos de parámetros 21, 22 a 2N que en el ejemplo mostrado están almacenados en una memoria 20 de la unidad de edición de voz 1. Asimismo, se muestran N sistemas 31, 32 a 3N que están unidos con la unidad de edición de voz 1 a través de un respectivo enlace de datos, tal como líneas individuales, un sistema de bus o canales de datos. Cada sistema puede realizar una edición de voz a través de la unidad de edición de voz. En particular, están presentes un ordenador 31 de a bordo con un juego de parámetros correspondiente para el ordenador 21 de a bordo, un sistema de navegación 32 con un juego de parámetros correspondiente para la navegación 22, un sistema de información de tráfico 33 con un juego de parámetros correspondiente para la información de tráfico 23, y un sistema de correo electrónico, tal como un sistema TTS 34, con un juego de parámetros correspondiente para el correo electrónico 24. Pueden preverse más sistemas 3N con un respectivo juego de parámetros asociado 2N. En el ejemplo mostrado es posible que, empleando una única unidad de edición de voz 1, el sistema de navegación 32 pueda hablar, por ejemplo, con una suave voz femenina que viene determinada por el juego de parámetros para el sistema de navegación 22. Asimismo, puede estar previsto, por ejemplo para avisos de tráfico, un juego de parámetros 23 con el que se emplee una fuerte voz de bajo masculino para la edición de voz.

La secuencia de ediciones de voz puede efectuarse sucesivamente en el tiempo con arreglo a la entrada del pedido de edición de voz desde los sistemas. Preferiblemente, las informaciones con mayor prioridad, por ejemplo informaciones del tráfico en situaciones de peligro, tal como falsos conductores, son las primeras en emitirse por cada edición de voz. De manera especialmente preferida, se emiten inmediatamente las informaciones con máxima prioridad, por ejemplo informaciones del ordenador de a bordo sobre funciones erróneas del vehículo o la iniciación de una calzada resbaladiza, pudiendo interrumpirse una edición de voz en curso. La edición de voz interrumpida puede llevarse después hasta su final o repetirse.

La invención tiene la ventaja de que sistemas con indicación acústica le proporcionen al conductor información de sistemas diferentes sin distraerle de su tarea, tal como ocurre en el caso de indicaciones visuales. Mediante la utilización de un dispositivo de síntesis de voz que se puede emplear por diferentes ordenadores de a bordo se pueden ahorrar costes. Se puede reducir la demanda de espacio de memoria en comparación con procedimientos de producción de voz empleados hasta ahora en, por ejemplo, sistemas de navegación.

La invención se puede utilizar de manera especialmente ventajosa en vehículos automóviles.

REIVINDICACIONES

- 5 1.- Dispositivo de edición diferenciada de voz (1) que puede unirse con un primer sistema (31) y al menos con otro sistema (32, 33 a 3N), estando asociada a la edición de voz del primer sistema (31) una primera característica de la voz y estando asociada a la edición de voz adicional del otro sistema (32, 33 a 3N) otra característica de la voz que se diferencia audiblemente de la primera característica de la voz, caracterizado por un dispositivo de síntesis de voz (10) que recibe parámetros de control que presentan una primera clase de parámetros dinámicos y una segunda clase de parámetros estáticos, controlando los parámetros dinámicos la articulación correspondiente al movimiento de un tracto vocal y controlando los parámetros estáticos las propiedades características de la voz, estando almacenados los parámetros estáticos para los sistemas como juegos de parámetros asociados en una memoria (20) del dispositivo de edición de voz y empleándose, en función de una señal de selección de un sistema, un juego de parámetros asociado por parte del dispositivo de síntesis de voz (10) para la edición de voz, y estando almacenados los parámetros dinámicos con arreglo a la secuencia de palabras, frases y series de frases en cada sistema.
- 10 2.- Dispositivo según la reivindicación 1, en el que los parámetros estáticos presentan una frecuencia fundamental de generador y/o formantes fijos que corresponden preferiblemente a la diferente dimensión geométrica del tracto vocal en un niño, una mujer o un hablante varón.
- 15 3.- Dispositivo según la reivindicación 2, en el que se pueden variar parámetros de generador y/o parámetros de formantes para la edición de voz de sistemas diferentes y se producen preferiblemente diferencias audibles en la prosodia, tales como la duración y/o la acentuación de segmentos de sílabas y/o la melodía de las frases.
- 20 4.- Dispositivo según cualquiera de las reivindicaciones 1 a 3, en el que el dispositivo de síntesis de voz (10) es un sintetizador de formantes con el que se puede influir sobre las propiedades características de la voz.
- 5 5.- Dispositivo según la reivindicación 4, en el que el sintetizador de formantes es adecuado para generar por separado sonidos con voz y sin voz, y en el que especialmente se pueden intercalar por medio de otros parámetros resonadores o miembros de amortiguación adicionales y/o se puede influir sobre los parámetros dinámicos para la articulación.
- 25 6.- Dispositivo según cualquiera de las reivindicaciones 1 a 5, en el que el dispositivo de síntesis de voz (10) está unido con un amplificador (12) y se efectúa una edición de voz a través de una salida de audio (14) del amplificador (12).
- 30 7.- Sistema para uso con un dispositivo según cualquiera de las reivindicaciones 1 a 6, con una primera salida para la emisión de parámetros dinámicos y una segunda salida para la emisión de una señal de selección para conmutar un juego de parámetros en el dispositivo de edición de voz (10).
- 8.- Sistema para uso con un dispositivo según cualquiera de las reivindicaciones 1 a 6, con una salida para la emisión de parámetros dinámicos y parámetros estáticos enviados preferiblemente como un juego de parámetros al dispositivo de edición de voz (10).
- 35 9.- Combinación de un dispositivo según cualquiera de las reivindicaciones 1 a 6 con al menos un primer sistema y un sistema adicional, tales como un ordenador (31) de a bordo, un sistema de navegación (32), un sistema de información de tráfico (33), un sistema de correo electrónico (34) o un sistema de información (3N), preferiblemente para uso en un vehículo.
- 10.- Procedimiento de edición diferenciada de voz empleando un dispositivo según cualquiera de las reivindicaciones 1 a 6.

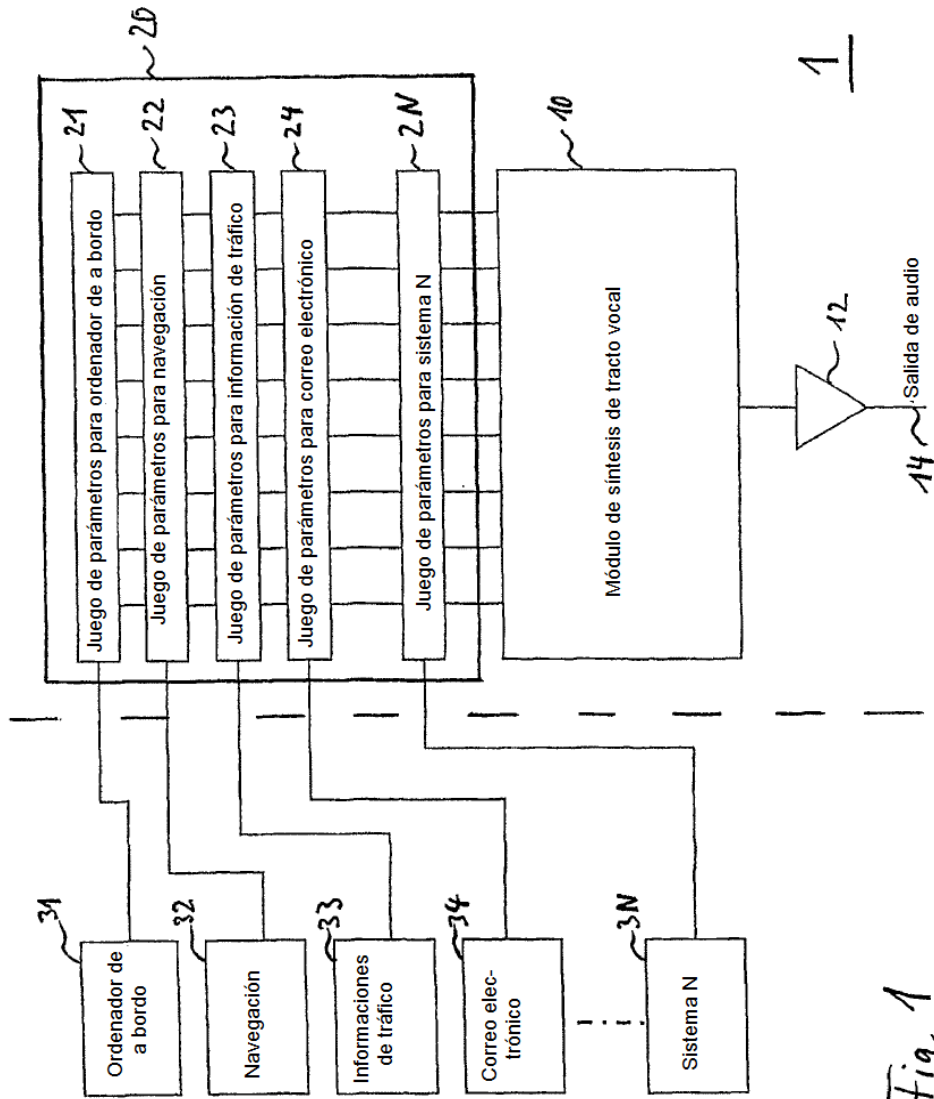


Fig. 1