

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 378 972**

51 Int. Cl.:

G10L 19/00 (2006.01)

G10L 11/04 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **07858612 .0**

96 Fecha de presentación: **17.10.2007**

97 Número de publicación de la solicitud: **2080194**

97 Fecha de publicación de la solicitud: **22.07.2009**

54 Título: **Atenuación de la sobresonorización, en particular para la generación de una excitación en un decodificador, en ausencia de información**

30 Prioridad:
20.10.2006 FR 0609225

73 Titular/es:
**FRANCE TELECOM
6 PLACE D'ALLERAY
75015 PARIS, FR**

45 Fecha de publicación de la mención BOPI:
19.04.2012

72 Inventor/es:
**VIRETTE, David y
KOVESI, Balazs**

45 Fecha de la publicación del folleto de la patente:
19.04.2012

74 Agente/Representante:
Linage González, Rafael

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).

ES 2 378 972 T3

DESCRIPCIÓN

Atenuación de la sobresonorización, en particular para la generación de una excitación en un decodificador, en ausencia de información.

5 La presente invención se refiere al procesamiento de señales digitales de audio, tales como señales de voz en telecomunicación, en particular a la descodificación de tales señales.

10 Se recuerda rápidamente que se puede predecir una señal de voz a partir de su pasado reciente (por ejemplo de 8 a 12 muestras a 8 kHz) mediante parámetros evaluados en ventanas cortas (10 a 20 ms en este ejemplo). Estos parámetros de predicción a corto plazo, representativos de la función de transferencia del conducto vocal (por ejemplo para pronunciar consonantes), se obtienen mediante procedimientos de análisis LPC (de "Linear Prediction Coding"). Se aplica también una correlación a más largo plazo para determinar periodicidades de sonidos sonoros (por ejemplo las vocales) debidas a la vibración de las cuerdas vocales. De este modo se trata de determinar al menos la frecuencia fundamental de la señal sonora que varía típicamente de 60 Hz (voz grave) a 600 Hz (voz aguda) según los locutores. Se determina entonces, por un análisis LTP (de "Long Term Prediction"), los parámetros LTP de un predictor a largo plazo, y en particular la inversa de la frecuencia fundamental, a menudo denominada "periodo de pitch". Se define entonces el número de muestras en una frecuencia fundamental por la relación F_e/F_0 (o su parte entera), donde:

- F_e es el ritmo de muestreo, y

- F_0 es la frecuencia fundamental.

20 Cabe resaltar que los parámetros de predicción a largo plazo LTP, como el periodo de pitch, representan la vibración fundamental de la señal de voz (cuando es sonora), mientras que los parámetros de predicción a corto plazo LPC representan la envoltura espectral de esta señal.

El conjunto de estos parámetros LPC y LTP, que resultan por lo tanto de una codificación de voz, se transmite por bloques hacia un decodificador homólogo, por una o varias redes de telecomunicación, para restituir a continuación la señal de voz inicial.

25 En el marco de la comunicación de tales señales por bloques, puede sobrevenir la pérdida de uno o más bloques consecutivos. Se entiende por el término "bloque" una sucesión de datos de señal que puede por ejemplo ser una trama en comunicación radio móvil, o también un paquete por ejemplo en comunicación sobre IP (de "Internet Protocol"), u otros.

30 En comunicación radio móvil por ejemplo, la mayoría de las técnicas de codificación por síntesis predictiva, y en particular la codificación de tipo CELP (de "Code Excited Liner Predictive"), proponen soluciones para recuperar tramas borradas. El decodificador es informado del caso de una trama borrada, por ejemplo por la transmisión de una información de borrado de trama procedente del decodificador canal. La recuperación de tramas borradas tiene por objetivo extrapolar los parámetros de la trama borrada a partir de una o más tramas anteriores consideradas como válidas. Algunos parámetros manipulados o codificados por los codificadores predictivos presentan una fuerte correlación entre tramas. Se trata típicamente de los parámetros de predicción a largo plazo LTP, para los sonidos sonoros por ejemplo, y de los parámetros de predicción a corto plazo LPC. Debido a esta correlación, es mucho más ventajoso reutilizar los parámetros de la última trama válida para sintetizar la trama borrada, que utilizar parámetros aleatorios, incluso erróneos.

En la generación de excitación CELP, los parámetros de la trama borrada se obtienen clásicamente como sigue.

40 Los parámetros LPC de una trama a reconstruir se obtienen a partir de los parámetros LPC de la última trama válida, por simple recopia de los parámetros o también con introducción de alguna atenuación (técnica utilizada por ejemplo en el codificador normalizado G723.1). A continuación, se detecta una sonorización o una no-sonorización en la señal de voz para determinar un grado de armonicidad de la señal al nivel de la trama borrada.

45 Si la señal es sorda, se puede generar una señal de excitación de manera aleatoria (por aparición de códigos de la excitación pasada, por una ligera atenuación de la ganancia de la excitación pasada, por selección aleatoria en la excitación pasada, o también utilizando códigos transmitidos que pueden ser totalmente erróneos).

50 Si la señal es sonora, el periodo de pitch (también denominada "retardo LTP") es generalmente la calculada para la trama anterior, eventualmente con una ligera "inestabilidad" (aumento del valor del retardo LTP para las tramas de error consecutivo, siendo tomada la ganancia LTP muy próxima a 1 o igual a 1). La señal de excitación se limita por lo tanto a la predicción a largo plazo efectuada a partir de una excitación pasada.

Los medios de disimulación de las tramas borradas, durante la descodificación, están generalmente muy ligados a la estructura del decodificador y pueden ser comunes a módulos de este decodificador, como por ejemplo el módulo de síntesis de la señal. Estos medios utilizan también señales intermedias disponibles en el seno del decodificador, como por ejemplo la señal de excitación pasada y memorizada durante el procesamiento de las tramas válidas que

preceden las tramas borradas.

Algunas técnicas utilizadas para disimular los errores producidos por paquetes perdidos durante el transporte de datos codificados según una codificación de tipo temporal recurren a menudo a técnicas de sustitución de formas de onda. Tales técnicas apuntan a reconstituir la señal seleccionando porciones de la señal descodificada antes del periodo perdido y no recurren a modelos de síntesis. Se aplican igualmente técnicas de suavización para evitar los fenómenos producidos por la concatenación de las diferentes señales.

Para los descodificadores que operan sobre señales codificadas mediante codificación por transformada, las técnicas de reconstrucción de las tramas borradas se apoyan generalmente en la estructura de codificación utilizada. Algunas técnicas apuntan a regenerar los coeficientes transformados perdidos a partir de los valores tomados por estos coeficientes antes del borrado.

Se han desarrollado otras técnicas de disimulación de las tramas borradas junto con la codificación canal. Sirven de informaciones proporcionadas por el descodificador canal, por ejemplo de informaciones relativas al grado de fiabilidad de los parámetros recibidos. Se indica en este caso que por el contrario, el objeto de la presente invención no presupone la existencia de un codificador canal.

Se ha propuesto, en el documento Combescure et al. :

"A 16, 24, 32 kbit/s Wideband Speech Codec Based on ATCELP", P.Combescure, J.Schnitzler, K.Ficher, R.Kirchherr, C.Lamblin, A.Le Guyader, D.Massaloux, C.Quinquis, J.Stegmann, P.Vary, Proceedings Conference ICASSP (1998), el uso de un procedimiento de disimulación de las tramas borradas equivalente al utilizado en los codificadores CELP para un codificador por transformada.

Los inconvenientes de este procedimiento eran la introducción de distorsiones espectrales audibles (voz "sintética", resonancias parásitas, etc.). Estos inconvenientes eran debidos en especial al uso de filtros de síntesis a largo plazo mal controlados (componente armónico único en sonidos sonoros, uso de porciones de la señal residual pasada en sonidos sordos). Además, el control de energía se efectúa aquí al nivel de la señal de excitación y la diana energética de esta señal se mantiene constante durante toda la duración del borrado, lo cual genera igualmente fenómenos audibles y molestos.

En el documento FR-2.813.722, se ha propuesto una técnica de disimulación de las tramas borradas, que no genera más distorsión a tasas de errores más elevadas y/o para intervalos borrados más largos. Esta técnica apunta a evitar el exceso de periodicidad para los sonidos sonoros y a controlar mejor la generación de la excitación no sonora. Para esto, se considera la señal de excitación (si es sonora) como la suma de dos señales:

- una componente fuertemente armónica limitada en banda a las frecuencias bajas del espectro total, y
- otra componente menos armónica y limitada a las frecuencias más altas.

La componente fuertemente armónica se obtiene por filtrado LTP. La segunda componente se obtiene también por un filtrado LTP convertido en no periódico por la modificación aleatoria de su periodo fundamental.

El problema principal de las técnicas de disimulación de error utilizadas hasta ahora en los codificadores CELP reside en la generación de la excitación sonora que, cuando se han perdido varias tramas consecutivas, pueden implicar un efecto de sobresonorización debido a la repetición de la misma frecuencia fundamental en varias tramas. De este modo se propone en el documento WO 2006/079348 hacer variar las muestras en las tramas sucesivas.

La presente invención tal como se define mediante las reivindicaciones 1, 7 y 8 mejoran la situación.

Se propone con este fin un procedimiento de síntesis de una señal digital de audio representado por bloques consecutivos de muestras, en el cual, tras la recepción de tal señal, para sustituir al menos un bloque inválido, se genera un bloque de sustitución a partir de las muestras de al menos un bloque válido, que precede el bloque inválido.

El procedimiento según la invención incluye las siguientes etapas:

- a) seleccionar un número elegido de muestras que forman una sucesión en al menos un último bloque válido que precede el bloque inválido,
- b) fragmentar la sucesión de muestras en grupos de muestras, y, en al menos una parte de los grupos, invertir muestras según reglas predeterminadas,
- c) reconcatenar los grupos cuyas muestras de algunos al menos se han invertido en la etapa b), para formar una parte al menos del bloque de sustitución, y
- d) si dicha parte obtenida en la etapa c) no llena todo el bloque de sustitución, copiar de nuevo dicha parte en el bloque de sustitución y aplicar de nuevo las etapas a), b), c) a dicha parte copiada de nuevo.

Esta inversión de muestras, que consiste por lo tanto en una manipulación de muestras muy simple y poco costosa en términos de cálculo y de medios de procesamiento, tiene por objetivo "romper" una sobrearmonicidad que pudiese haber estado presente si se hubiese aplicado una simple recopia de periodo de pitch.

De este modo, entre las ventajas que ofrece la presente invención, su aplicación solo necesita un coste de cálculo muy bajo.

La invención se aplica ventajosamente al caso en el que la señal digital de audio es una señal de voz sonora, y, más particularmente, escasamente sonora ya que la simple recopia de periodo de pitch da resultados mediocres en este caso. De este modo, según una característica ventajosa, se detecta un grado de sonorización en la señal de voz y se aplican las etapas a) a d) si la señal es al menos escasamente sonora.

La presente invención se apoya ventajosamente en la frecuencia fundamental de la señal digital de audio para constituir los grupos en la etapa b). De este modo, ventajosamente, en la etapa a):

a1) se detecta un tono en la señal digital de audio, y

a2) dicho número elegido de muestras seleccionadas en la etapa a) corresponde al número de muestras que incluye un periodo que corresponde a la inversa de una frecuencia fundamental del tono detectado.

Por supuesto, en el caso de una señal de voz, la operación a1) puede consistir en detectar una sonorización y la operación a2) apuntaría, si la señal de voz es sonora, a seleccionar un número de muestras que se extiende sobre todo uno periodo de pitch (inversa de una frecuencia fundamental de un tono de voz). Sin embargo, se destaca que esta realización también puede apuntar a una señal distinta de una señal de voz, en particular una señal musical, si se puede detectar en la misma una frecuencia fundamental propia de un tono global de música.

En una realización, la fragmentación de la etapa b) se efectúa por grupos de dos muestras, y se invierten las posiciones de las muestras de un mismo grupo la una respecto de la otra.

Sin embargo, en esta realización, cabe distinguir el caso en el que el periodo de pitch (o más generalmente el periodo inverso de la frecuencia fundamental) incluye un número de muestras par o impar. En particular, si el número de muestras que incluye el periodo del tono detectado es un número par, un número impar de muestras (preferentemente una sola muestra) se añade ventajosamente o se sustrae de las muestras de dicho periodo para formar la selección de la etapa a).

Conviene señalar también lo que se entiende por las "reglas predeterminadas de inversión". Estas reglas, que se pueden elegir según las características de la señal recibida, imponen en particular el número de muestras por grupos en la etapa b) y la manera de invertir las muestras en un grupo. En la realización anterior, se prevén grupos de dos muestras y una simple inversión de las posiciones respectivas de estas dos muestras. Sin embargo, otras configuraciones son posibles (grupos que incluyen más de dos muestras y permutación de todas las muestras de tales grupos). Por otra parte, las reglas de inversión pueden fijar también el número de grupos en los cuales la inversión se efectúa. Una realización particular consiste en convertir en aleatorios los casos de inversión de muestras en cada grupo y fijar un umbral de probabilidad para invertir o no las muestras de un grupo. Este umbral de probabilidad puede tener un valor fijo, o también un valor variable y depender ventajosamente de una función de correlación relativa al periodo de pitch. En este caso, la propia determinación formal del periodo de pitch no es necesaria. Por otra parte, más generalmente, el procesamiento en el sentido de la invención se puede efectuar también si la señal válida recibida no es simplemente sonora, en cuyo caso, no existe realmente periodo de pitch detectable. En este caso, se puede prever la fijación de un número dado de muestras arbitrario (por ejemplo doscientas muestras) y realizar el procesamiento en el sentido de la invención sobre este número de muestras. También es posible tomar el valor correspondiente al máximo de la función de correlación limitando la búsqueda a un intervalo de valor (por ejemplo entre $MAX_PITCH/2$ y MAX_PITCH , donde MAX_PITCH es el valor máximo en la búsqueda de periodo de pitch).

La presente invención, al proponer de este modo la atenuación de sobresonorización, ofrece las siguientes ventajas:

- la voz sintetizada durante una pérdida de bloque ya no presenta prácticamente del fenómeno de sobrearmonicidad o de sobresonorización, y

- la complejidad necesaria para generar una excitación sonora es muy reducida, como se verá en el ejemplo de realización descrito en detalle en lo sucesivo

Por otra parte, otras ventajas u características de la invención se harán evidentes tras examinar la descripción detallada, dada a título de ejemplo en lo sucesivo, y de los dibujos anexos en los cuales:

- la figura 1 ilustra el principio de una generación de excitación que permite atenuar el efecto de sobresonorización, integrando una inversión aleatoria de muestras, sobre bloques de dos muestras y con una probabilidad del 50% en el ejemplo representado, sobre todo un periodo de pitch,

- la figura 2 ilustra el principio de una generación de excitación que integra una inversión de muestras, aquí

sistemática, sobre bloques de dos muestras en el ejemplo representado y sobre todo un periodo de pitch,

- la figura 3a ilustra la aplicación de la inversión sistemática de la figura 2 sen una señal de la cual se ha estimado un periodo de pitch que incluye un número de muestras impar,

5 - la figura 3b representa, a título meramente ilustrativo, la aplicación de la inversión sistemática de la figura 2 sobre una señal de la cual se ha estimado un periodo de pitch que incluye un número de muestras par,

- la figura 3c ilustre la aplicación de la inversión sistemática de la figura 2, con aquí una corrección por adición de una muestra a la duración correspondiente al periodo de pitch, para convertir esta duración impar en término de número de muestras que incluye,

10 - la figura 4 ilustra esquemáticamente las etapas principales de un procedimiento en el sentido de la invención, a la decodificación,

- la figura 5 ilustra muy esquemáticamente la estructura de un aparato de recepción de una señal digital de audio que incluye un dispositivo de síntesis para la aplicación del procedimiento en el sentido de la invención.

15 En primer lugar se hace referencia a la figura 4 para ilustrar el contexto de aplicación de la presente invención. Al recibir una señal de entrada Se durante la decodificación, se detecta (prueba 50) la pérdida de uno o más bloques consecutivos. Si no se constata ninguna pérdida de bloque (flecha O al concluir la prueba 50), evidentemente no se plantea ningún problema, y se acaba el procesamiento de la figura 4.

Por el contrario, se constata la pérdida de uno o más bloques consecutivos (flecha N al concluir la prueba 50), se detecta entonces el grado de sonorización (prueba 51) de la señal.

20 Si la señal no es sonora (flecha N al concluir la prueba 51), se sustituyen los bloques perdidos por ejemplo por un ruido blanco, audible, denominado "ruido de confort" 52, y se añade la ganancia 61 de las muestras de los bloques así reconstruidos. Se puede realizar por ejemplo un control sobre la energía de la señal reconstruida Ss, con adaptación de la ley de evolución, y/o hacer evolucionar parámetros del modelo hacia una señal de reposo tal como el ruido de confort 52.

25 En una variante de la presente invención, solo se consideran dos clases de señales, las señales sonoras por una parte, y las señales escasamente sonoras o sordas por otra parte. La ventaja de esta variante es que la generación de la señal sorda será idéntica a la síntesis escasamente sonora. Como se ha indicado anteriormente, el "periodo de pitch" utilizado para las señales sordas es un valor aleatorio, preferiblemente bastante grande (por ejemplo doscientas muestras). En un bloque sordo, la señal anterior es no-armónica, aplicando el procesamiento en el sentido de la invención durante un periodo suficientemente grande, se garantiza que la señal así generada permanece no-armónica. La naturaleza de la señal se conservará ventajosamente, lo cual no sería el caso si se utilizase una señal generada aleatoriamente (por ejemplo un ruido blanco).

30 Si la señal es fuertemente sonora (flecha O al concluir la prueba 51), se sustituyen los bloques perdidos por recopia del periodo de pitch T. Se determina entonces el periodo de pitch T identificado en la última parte todavía válida de la señal recibida Se (por cualquier técnica 53 conocida en sí). Se copian de nuevo a continuación las muestras de este periodo de pitch T en los bloques perdidos (referencia 54). Se aplica a continuación una ganancia apropiada 61 a las muestras así sustituidas (para efectuar por ejemplo una atenuación o "fading").

En el ejemplo descrito, si la señal es medianamente sonora (o, en una variante menos sofisticada pero más general, si la señal es simplemente sonora), se aplica el procedimiento en el sentido de la invención (flecha M al concluir la prueba 51 en el grado de sonorización).

40 Con referencia a las figuras 1 y 2, el principio de la invención consiste en unir las muestras de los últimos bloques válidos recibidos, por grupos de al menos dos muestras. En el ejemplo de las figures 1 y 2, se han agrupado efectivamente estas muestras de dos en dos. Sin embargo, se podrían agruparlas en cantidades superiores a de dos en dos, en cuyo caso, las reglas de inversión de muestras por grupo y de toma en consideración de la paridad en número de muestras del periodo de pitch T, descritas en detalle en lo sucesivo, estarían ligeramente adaptadas.

45 Con referencia en particular a la figura 2, los grupos A, B, C, D, de dos muestras en los últimos bloques válidos recibidos se vuelven a copiar y se concatenan a las últimas muestras recibidas. Sin embargo, en estos grupos recopiados, referenciados A', B', C', D', se han invertido los valores de las dos muestras en cada grupo (o conservado su valor e invertido sus posiciones respectivas). De este modo, el grupo A se convierte en el grupo A', con sus dos muestras invertidas respecto del grupo A (según dos flechas del grupo A' en la figura 2). El grupo B se convierte en el grupo B', con sus dos muestras invertidas respecto del grupo B, y así sucesivamente. La recopia y la concatenación de los grupos A', B', C', D', se efectúa ventajosamente respetando el periodo de pitch T. De este modo, el grupo A', constituido por las muestras invertidas del grupo A, se separa del grupo A en un número de muestras que corresponde a la duración del periodo de pitch T. Asimismo, el grupo B' se separa del grupo B por una duración que corresponde al periodo de pitch T, y así sucesivamente.

En la figura 2, la inversión de las muestras por grupo es sistemática. En una variante tal como la representada en la figura 1, se puede hacer que sea aleatorio el caso de esta inversión. Se puede incluso prever la fijación de un umbral p de probabilidad para invertir o no las muestras de un grupo. En el ejemplo representado en la figura 1, el umbral p se fija en el 50% de manera que solo dos grupos B', C', de cuatro, tienen sus muestras invertidas. Se puede asimismo prever hacer que sea variable el umbral p de probabilidad, en particular hacerle depender de una función de correlación relativa al periodo de pitch T, como se verá más adelante.

Retomando la descripción de la realización ilustrada en la figura 2, donde se aplica una inversión sistemática de las muestras por grupo, se obtiene, con referencia ahora a la figura 3a, una nueva sucesión de muestras T', de duración correspondiente al periodo de pitch T, pero con inversión de las muestras de dos en dos. Se han representado en la figura 3a las últimas muestras de los últimos bloques válidos recibidos en la señal Se y que se han memorizado en un descodificador. Aquí, como la inversión es sistemática y no aleatoria con estimación de una correlación, se ha determinado el periodo de pitch T de la señal sonora (por un medio conocido en sí) y se han recogido las últimas muestras 10, 11, ..., 22 de la señal Se, que se extienden a lo largo de la duración del periodo de pitch T. Las dos primeras muestras 10 y 11 se invierten en la señal a reconstruir, referenciada Ss. Las tercera y cuarta muestras 12 y 13 se invierten también, y así sucesivamente. Se obtiene entonces una sucesión T' de muestras 11, 10, 13, 12... que se extiende a lo largo de una misma duración que la frecuencia fundamental. Si varios bloques que se extienden a lo largo de varios periodos de pitch carecen de descodificación, se sigue la reconstrucción de la señal Ss tomando la sucesión T' y reiniciando la inversión de las muestras de dos en dos de la sucesión T', para obtener una nueva sucesión T'', y así sucesivamente.

En el caso de la figura 3a, el número de muestras por periodos T, T', T'' es igual a un mismo número impar (trece muestras en el ejemplo representado), lo cual permite obtener una mezcla progresiva de las muestras a medida que avanza la reconstrucción de la señal Ss, y de ahí, una atenuación eficaz de la sobrearmonicidad (o, dicho de otro modo, de la sobreesonorización de la señal reconstruida).

Por el contrario, en el caso ilustrado en la figura 3b donde el número de muestras por periodos T, T', T'' es un número par (doce muestras en el ejemplo representado), practicando dos veces una inversión (del periodo T al periodo T', y después del periodo T' al periodo T'') de las muestras, tomadas de dos en dos, del periodo de pitch T, se encuentra exactamente la misma sucesión que el periodo de pitch T en la sucesión T'', lo cual genera entonces una sobrearmonicidad.

Este problema se puede solucionar modificando el número de muestras a invertir por grupo (y tomar por ejemplo un número impar de muestras por grupo).

Sin embargo, se ha ilustrado otra realización en la figura 3c. Esta realización consiste simplemente, cuando el periodo de pitch incluye un número par de muestras y cuando las inversiones apuntan a números pares de muestras por grupo, en añadir un número impar de muestras a la frecuencia fundamental de la señal a reconstruir. En la figura 3c, el último periodo de pitch detectado T incluye doce muestras 31, 32, ..., 42. Se añade entonces una muestra al periodo de pitch y se obtiene un periodo T+1 que incluye un número impar de muestras. De este modo, en el ejemplo ilustrado en la figura 3c, la muestra 30 se convierte en la primera muestra de la memoria a partir de la cual se aplica la inversión de muestras de dos en dos como se ha ilustrado en la figura 2 (o la figura 3a). Se obtiene un periodo T' de la señal reconstruida Ss, que incluye un número de muestras impar, a la cual se aplica todavía la inversión de muestras de dos en dos para obtener el periodo T'', que incluye todavía un número de muestras impar, y así sucesivamente. Se observará entonces que la sucesión de muestras 33, 30, 35, 32, 34... del periodo T'' es muy diferente, esta vez, de la sucesión de muestras 30, 31, 32, 33... del periodo de pitch inicial T.

Con referencia de nuevo a la figura 4 que aplica, en el ejemplo representado, la realización ilustrada en las figuras 2, 3a y 3c, cuando la señal Se es medianamente sonora (flecha M al concluir la prueba 51), se determina el periodo de pitch T en las últimas muestras de la señal Se válidamente recibidas (por una técnica 56 que puede ser conocida en sí). Se detecta si el número de muestras en el periodo de pitch T es par o impar. Si este número es impar (flecha N al concluir la prueba 57), se aplica directamente la inversión de las muestras de dos en dos (etapa 58) como se describe en lo sucesivo con referencia a la figura 3a. Si el número de muestras en el periodo de pitch T es par (flecha O al concluir la prueba 57), se añade una muestra al periodo de pitch T (etapa 59) y se aplica a continuación la inversión de las muestras de dos en dos (etapa 58), según el procesamiento descrito anteriormente con referencia a la figura 3c. A continuación, se aplica eventualmente una ganancia elegida 61 a la sucesión de muestras así obtenida para formar la señal finalmente reconstruida Ss.

Como se ha indicado anteriormente con referencia a la figura 4, el periodo de pitch se calcula en primer lugar a partir de una o algunas tramas anteriores. A continuación, la excitación de armonicidad reducida se genera de la manera ilustrada en la figura 2, con inversión sistemática. Sin embargo, en la variante ilustrada en la figura 1, se puede generar con inversión aleatoria. Esta inversión irregular de las muestras de la excitación sonora permite ventajosamente atenuar la sobre-armonicidad. Se detalla en lo sucesivo esta realización ventajosa.

Habitualmente, en forma de simple recopia de periodo de pitch, la excitación sonora se calcula según una fórmula del tipo:

$$s(n) = g_{lp} \cdot s(n-T) \quad (1)$$

donde T es el periodo de pitch estimado y g_{lp} es una ganancia LTP elegida.

En una realización de la invención, la excitación sonora se calcula por grupo de dos muestras y con inversión aleatoria según el siguiente procesamiento.

- 5 En primer lugar, se genera un número aleatorio x en el intervalo [0; 1]. A continuación, en función del valor de x:

- si $x < p$, s(n) y s(n+1) se calculan a partir de la ecuación (1)
- si $x \geq p$, s(n) y s(n+1) se calculan según las siguientes ecuaciones (2) y (3):

$$s(n) = g_{lp} \cdot s(n-T+1) \quad (2)$$

$$s(n+1) = g_{lp} \cdot s(n-T) \quad (3)$$

- 10 El valor p representa la probabilidad de invertir las dos muestras s(n) y s(n+1). Por ejemplo, se puede fijar el valor p de manera que $p = 50\%$.

En una variante ventajosa, se puede también elegir una probabilidad variable, por ejemplo de la forma:

$$p = \text{corr} \quad (4)$$

- 15 donde la variable corr corresponde al valor máximo de la función de correlación en el periodo de pitch, referenciada Corr(T). Para un periodo de pitch T, la función de correlación Corr(T) se calcula utilizando solo $2 \cdot T_m$ muestras al final de la señal memorizada, y:

$$\text{Corr}(T) = \frac{2 \sum_{i=Lmem-2T_m+T}^{Lmem-1} m_i m_{i-T}}{\sum_{i=Lmem-2T_m}^{Lmem-1} m_i^2 + \sum_{i=Lmem-2T_m+T}^{Lmem-1-T} m_i^2} \quad (5)$$

donde $m_0 \dots m_{Lmem-1}$ son las últimas muestras de la señal descodificada anteriormente, y siguen todavía disponibles en la memoria del descodificador.

- 20 A partir de esta fórmula, se entenderá que la longitud de esta memoria L_{mem} (en número de muestras almacenadas) debe ser igual a al menos dos veces el valor máximo de la duración del periodo de pitch (en número de muestras). Para tener en cuenta las voces más graves (de frecuencia fundamental más baja del orden de 50 Hz), el número de muestras a almacenar puede ser del orden de 300, para una tasa reducida de muestreo en banda estrecha, y de más de 300 para tasas de muestreo más elevadas.

- 25 La función de correlación corr(T), dada por la fórmula (5), alcanza un valor máximo cuando la variable T corresponde al periodo de pitch T_0 y este valor máximo da una indicación del grado de sonorización. Típicamente, si este valor máximo es muy próximo a 1, entonces la señal es fuertemente sonora. Si está próximo a 0, la señal no es sonora.

- Se entenderá de este modo que en esta realización, la determinación previa del periodo de pitch no es necesaria para construir los grupos de muestras a invertir. En particular, la determinación del periodo de pitch T_0 se puede efectuar junto con la constitución de los grupos en el sentido de la invención, por aplicación de la fórmula (5) anterior.

Si la señal es muy sonora, entonces la probabilidad p será muy grande, y la sonorización se conservará conforme el

cálculo según la fórmula (1). Si, por el contrario, la sonorización de la señal Se no es muy pronunciada, la probabilidad p será menor se utilizarán ventajosamente las ecuaciones (2) y (3).

Por supuesto, también se pueden utilizar otros cálculos de correlaciones.

5 Por ejemplo, también es posible calcular la excitación armónica en función de clases predefinidas. Para las clases muy sonoras, se utilizará más bien la ecuación (1). Para las clases mediana o escasamente sonoras, se utilizarán más bien las ecuaciones (2) y (3). Para las clases sordas, no se genera ninguna excitación armónica y la excitación se puede entonces generar a partir de un ruido blanco. Sin embargo, en la variante descrita anteriormente, e utilizarán también las ecuaciones (2) y (3) con un periodo de pitch arbitrario suficientemente grande.

10 De manera más general, la presente invención no se limita a las realizaciones descritas anteriormente a título de ejemplo; se extiende a otras variantes.

15 En el contexto de realización de la invención descrita en detalle anteriormente, la generación de excitación en codificación por síntesis predictiva CELP apunta a evitar la sobresonorización en el contexto de la disimulación de errores de transmisiones de tramas. Se puede prever sin embargo la utilización de los principios de la invención para extensión de banda. Se puede entonces utilizar la generación de una excitación en banda ensanchada en un sistema de extensión de banda (con o sin transmisión de informaciones), basada en un modelo de tipo CELP (o subbanda CELP). La excitación de la banda alta se puede calcular entonces como se ha descrito anteriormente, lo cual permite limitar entonces la sobrearmonicidad de esta excitación.

20 Por otra parte, la aplicación de la invención está particularmente adaptada a la transmisión de señales en redes por tramas, o también por paquetes, por ejemplo paquetes de "voz sobre IP" (de "Internet Protocol"), de manera a proporcionar una calidad aceptable durante la pérdida de tales paquetes sobre IP, y garantizando, sin embargo, al mismo tiempo una complejidad limitada.

Por supuesto, la inversión de las muestras se puede llevar a cabo en grupos de muestras de dimensión superior a dos.

25 Por otra parte, se ha descrito anteriormente la generación de un bloque de sustitución de un bloque inválido a partir de las muestras de un bloque válido, que precede el bloque inválido. En una variante, se puede usar más bien un bloque válido que sucede al bloque inválido para realizar la síntesis del bloque inválido (síntesis a posteriori). Esta realización puede ser ventajosa en particular para sintetizar varios bloques inválidos sucesivos y, en particular, para sintetizar:

30 - bloques inválidos que suceden inmediatamente a bloques válidos anteriores, a partir de estos bloques anteriores, y a continuación

- bloques inválidos que preceden inmediatamente a los siguientes bloques válidos, a partir de estos siguientes bloques.

35 La presente invención apunta también a un programa informático destinado a ser almacenado en la memoria de un dispositivo de síntesis de una señal digital de audio. Este programa incluye entonces instrucciones para la aplicación del procedimiento en el sentido de la invención, cuando es ejecutado por un procesador de tal dispositivo de síntesis. Por otra parte, la figura 4 descrita anteriormente puede ilustrar un organigrama de tal programa informático.

Por otra parte, la presente invención apunta también a un dispositivo de síntesis de una señal digital de audio constituida por una sucesión de bloques. Este dispositivo podría incluir por otra parte una memoria que almacena el programa informático mencionado anteriormente. Con referencia a la figura 5, este dispositivo SYN, incluye:

40 - una entrada E para recibir bloques de la señal Se , que precede a al menos un bloque actual a sintetizar, y

- una salida S para proporcionar la señal sintetizada Ss y que incluye al menos este bloque actual a sintetizar.

45 El dispositivo de síntesis SYN en el sentido de la invención incluye medios tales como una memoria de trabajo MEM (o de almacenamiento del programa informático mencionado anteriormente) y un procesador PROC que coopera con esta memoria MEM, para la aplicación del procedimiento en el sentido de la invención, y para sintetizar de este modo el bloque actual a partir de al menos uno de los bloques anteriores de la señal Se .

La presente invención apunta también a un aparato de recepción de una señal digital de audio constituida por una sucesión de bloques, tal como un decodificador de tal señal por ejemplo. Con referencia también a la figura 5, este aparato puede incluir ventajosamente un detector de bloques inválidos DET, así como el dispositivo SYN en el sentido de la invención para sintetizar bloques inválidos detectados por el detector DET.

REIVINDICACIONES

1. Procedimiento de síntesis de una señal digital de audio representada por bloques consecutivos de muestras, en el cual, tras la recepción de tal señal, para sustituir al menos un bloque inválido, se genera un bloque de sustitución a partir de las muestras de al menos un bloque válido que precede al bloque inválido, caracterizado porque incluye las siguientes etapas:
 - 5 a) estimar una correlación que permite detectar un eventual periodo correspondiente a la inversa de una frecuencia fundamental de un tono en la señal digital de audio, y seleccionar un número (T) de muestras que es función de esta estimación, que forman una sucesión en al menos un último bloque válido que precede al bloque inválido,
 - b) fragmentar la sucesión de muestras en grupos (A, B, C, D) de dos muestras, y, en al menos una parte de los grupos, invertir las dos muestras o no en función de dicha estimación de correlación,
 - 10 c) reconcatenar los grupos (A', B', C', D') cuyas muestras de algunos al menos se han invertido en la etapa b), para formar una parte (T') al menos del bloque de sustitución, y
 - d) si dicha parte obtenida en la etapa c) no llena todo el bloque de sustitución, copiar de nuevo dicha parte (T') en el bloque de sustitución y aplicar de nuevo las etapas b) y c) a dicha parte copiada.
2. Procedimiento según la reivindicación 1, en el cual la señal digital de audio es una señal de voz, caracterizado porque la estimación de correlación incluye la detección de un grado de sonorización (51) en la señal de voz y se aplican las etapas b) a d) si la señal es escasamente sonora o sorda.
- 15 3. Procedimiento según una de las reivindicaciones anteriores, caracterizado porque, para llevar a cabo la etapa a):
 - a1) se busca una correlación para detectar un eventual periodo correspondiente a la inversa de una frecuencia fundamental de un tono en la señal digital de audio (56), y
 - 20 a2) dicho número de muestras seleccionadas en la etapa a) corresponde:
 - al número de muestras que incluye un periodo correspondiente a la inversa de una frecuencia fundamental del tono si la búsqueda de correlación detecta dicho periodo, y
 - si no, a un número fijo de muestras, predeterminado.
4. Procedimiento según la reivindicación 3, caracterizado porque, si el número de muestras que incluye el periodo del tono detectado es un número par, un número impar de muestras (30) se añade a o se sustrae de las muestras de dicho periodo para formar la selección de la etapa a), no estando una muestra de la sucesión (T') así formada en la etapa a) agrupada, en la etapa b), con otra muestra de dicha sucesión (T').
- 25 5. Procedimiento según una de las reivindicaciones anteriores, caracterizado porque dichas reglas predeterminadas imponen convertir en aleatorios los casos de inversión de muestras en cada grupo y fijan un umbral de probabilidad (p) para invertir o no las muestras de un grupo.
- 30 6. Procedimiento según la reivindicación 5, caracterizado porque el umbral de probabilidad (p) es variable y depende de la estimación de correlación.
7. Programa informático destinado a ser almacenado en la memoria de un dispositivo de síntesis de una señal digital de audio, caracterizado porque incluye instrucciones para la aplicación del procedimiento según una de las reivindicaciones 1 a 6 cuando es ejecutado por un procesador de tal dispositivo de síntesis.
- 35 8. Dispositivo de síntesis de una señal digital de audio constituida por una sucesión de bloques, que incluye:
 - una entrada para recibir bloques de la señal (Se), que precede a al menos un bloque actual a sintetizar, y
 - una salida para proporcionar la señal sintetizada (Ss) y que incluye al menos dicho bloque actual,caracterizado porque incluye medios (MEM, PROC) adaptados para la aplicación del procedimiento según una de las reivindicaciones 1 a 6, para sintetizar el bloque actual a partir de al menos uno de dichos bloques precedentes.
- 40 9. Aparato de recepción de una señal digital de audio constituida por una sucesión de bloques, que incluye un detector de bloques inválidos (DET), caracterizado porque incluye, además, un dispositivo (SYN) según la reivindicación 8, para sintetizar bloques de sustitución de los bloques inválidos.

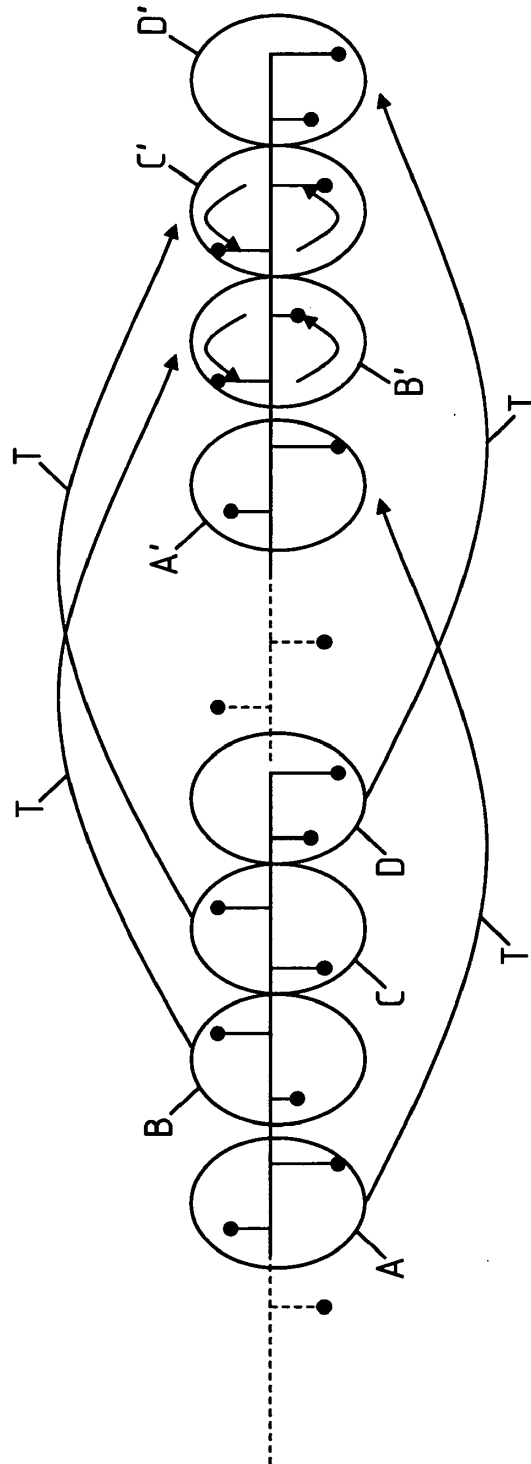


FIG. 1

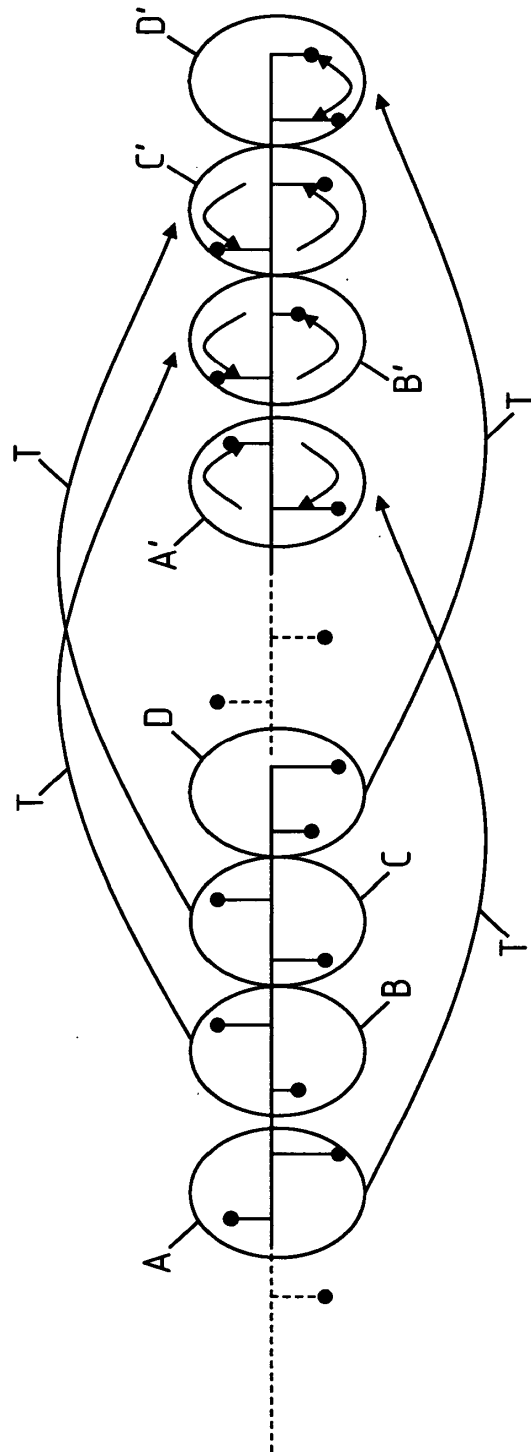


FIG. 2

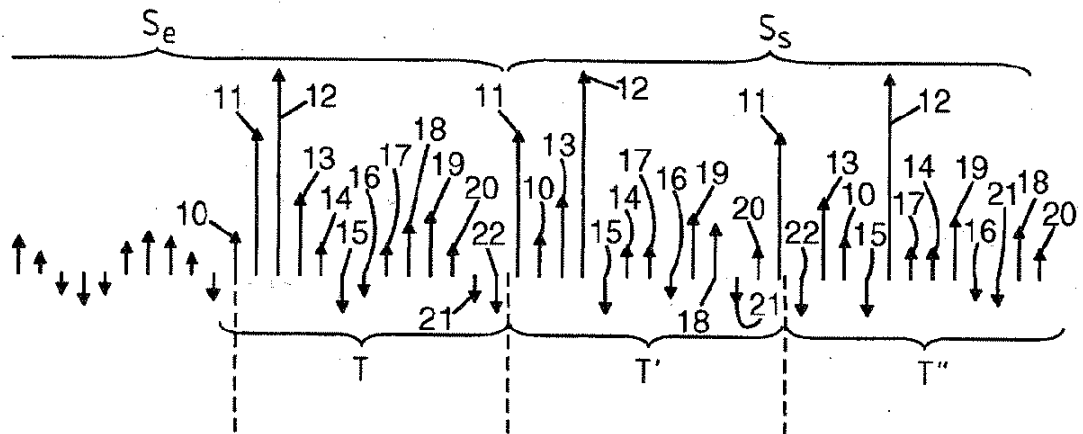


FIG. 3a

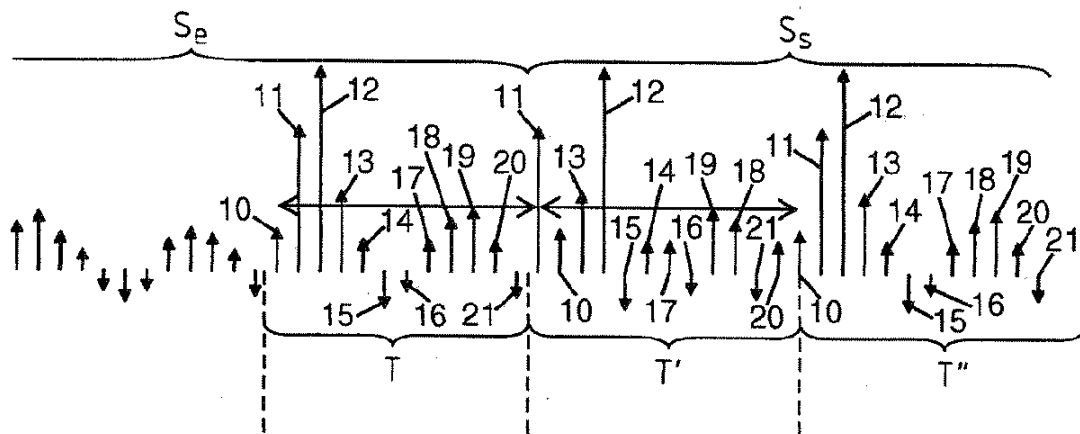


FIG. 3b

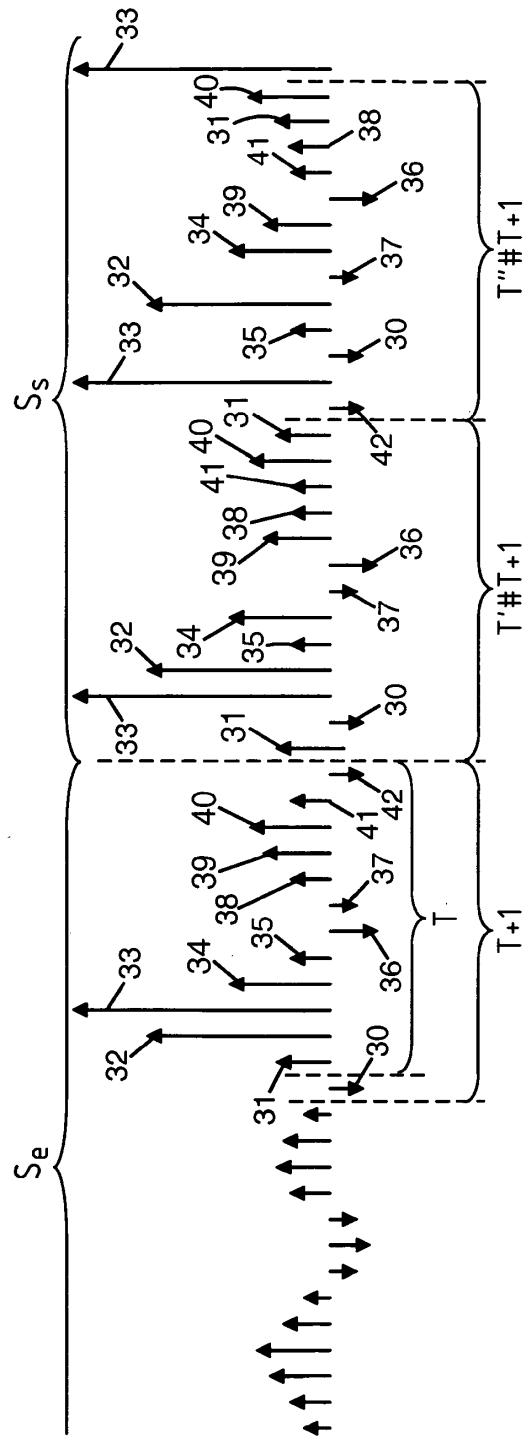


FIG. 3c

FIG. 4

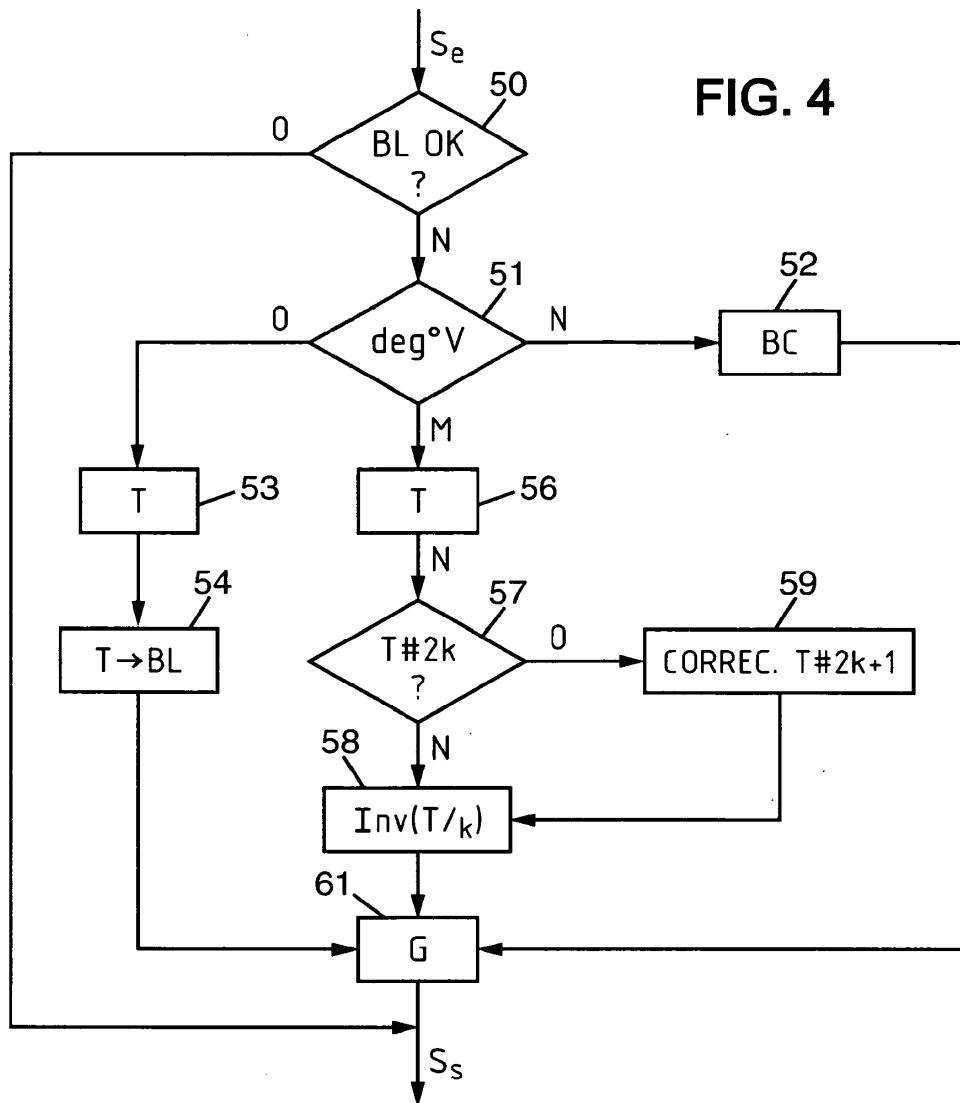


FIG. 5

