(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 294 283**

(51) Int. Cl.:
***G10L 15/28*** (2006.01)
***G10L 15/14*** (2006.01)
***G10L 15/18*** (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Número de solicitud europea: **03727600 .3**
(86) Fecha de presentación : **20.03.2003**
(87) Número de publicación de la solicitud: **1490863**
(87) Fecha de publicación de la solicitud: **29.12.2004**

(54) Título: **Procedimiento de reconocimiento de voz por medio de un solo transductor.**

(30) Prioridad: **29.03.2002 FR 02 04286**

(45) Fecha de publicación de la mención BOPI:
**01.04.2008**

(45) Fecha de la publicación del folleto de la patente:
**01.04.2008**

(73) Titular/es: **FRANCE TELECOM**
**6, place d'Alleray**
**75015 Paris, FR**

(72) Inventor/es: **Ferrieux, Alexandre y**
**Delphin-Poulat, Lionel**

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

## DESCRIPCIÓN

Procedimiento de reconocimiento de voz por medio de un solo transductor.

La presente invención se refiere a un procedimiento de traducción de datos de entrada en al menos una secuencia lexical de salida, que incluye una etapa de decodificado de los datos de entrada en el transcurso de la cual entidades lexicales de las cuales los mencionados datos son representativas se identifican por medio de al menos un modelo.

Tales procedimientos se utilizan corrientemente en aplicaciones de reconocimiento de voz, donde se utiliza al menos un modelo para reconocer informaciones presentes en los datos de entrada, pudiendo una información estar constituida por ejemplo por un conjunto de vectores de parámetros de un espacio acústico continuo, o también por una etiqueta atribuida a una entidad sublexical.

La solicitud de patente EP 0715 298 describe un procedimiento de reconocimiento de voz que incluye dos etapas distintas de decodificado que son respectivamente realizadas por medio de modelos fonéticos y lexicales. En el transcurso de una etapa, se definen fronteras temporales entre diferentes fonemas y se establece en el transcurso de otra etapa una lista de diferentes fonemas identificados como que pueden ser incluidos dentro de cada intervalo así definidos, con probabilidades asociadas. Esta identificación se realiza por comparación con un modelo acústico.

La publicación de RAMANUJAN *et al*. titulada "Address code and arithmetic optimization for embedded systems" DESIGN AUTOMATION CONFERENCE, 2002 publica un método que permite optimizar el direccionado de una memoria y en particular racionalizar las atribuciones de direcciones a grandes conjuntos de datos. Esta publicación prevé, después de haber observado que la mayoría de los datos de un mismo conjunto son en general parecidos los unos a los otros cuando se almacenan en memoria, deducir una nueva dirección escalar de un dato de un conjunto particular a partir de una dirección escalar ya conocida de otro dato que pertenece a este mismo conjunto, más bien que desarrollar todo un nuevo proceso de cálculo similar al que ha permitido obtener la dirección ya conocida.

En algunas aplicaciones, el calificativo "lexical" se aplicará a una frase considerada en su conjunto, como secuencia de palabras, y las entidades sublexicales, serán entonces palabras, mientras que en otras aplicaciones, el calificativo "lexical" se aplicará a una palabra, y las entidades sublexicales serán entonces fonemas o también sílabas aptas para formar tales palabras, si estas son de naturaleza literal, o cifras, si las palabras son de naturaleza numérica, es decir números.

Una primera aproximación para realizar un reconocimiento de voz consiste en utilizar un tipo particular de modelo que presente una topología regular y esté destinado para aprender todas las variantes de pronunciación de cada entidad lexical, es decir por ejemplo una palabra, incluida en el modelo. Según esta primera aproximación, los parámetros de un conjunto de vectores acústicos adecuado para cada información que está presente en los datos de entrada y corresponde a una palabra desconocida deben ser comparados con conjuntos de parámetros acústicos que corresponden cada uno a uno de los muy numerosos símbolos contenidos en el modelo, con el fin de identificar un símbolo modelizado al cual corresponde lo más verosímilmente esta información. Una aproximación de este tipo garantiza en teoría un alto porcentaje de conocimiento si el modelo utilizado está bien concebido, es decir casi-exhaustivo, pero una casi-exhaustividad de este tipo solo puede ser obtenida a costa de un largo proceso de aprendizaje del modelo, que debe asimilar una enorme cantidad de datos representativos de todas las variantes de pronunciación de cada una de las palabras incluidas en este modelo. Este aprendizaje se realiza en principio haciendo pronunciar por un gran número de personas todas las palabras de un vocabulario dado, y registrando todas las variantes de pronunciación de estas palabras. Resulta claramente que la construcción de un modelo lexical casi-exhaustivo no se puede considerar en la práctica para vocabularios que presentan un tamaño superior a algunos centenares de palabras.

Una segunda aproximación ha sido concebida con el fin de reducir el tiempo de aprendizaje necesario para las aplicaciones de reconocimiento de voz, reducción que es esencial para aplicaciones de traducción en vocabularios muy grandes que pueden contener varios centenares de millares de palabras, cuyo segundo acercamiento consiste en realizar una descomposición de las entidades lexicales considerándolas como conjuntos de entidades sublexicales, en utilizar un modelo sublexical que modelice las indicadas entidades sublexicales con miras a permitir su identificación en los datos de entrada, y un modelo de articulación que modelice diferentes combinaciones posibles de estas entidades sublexicales.

Una aproximación de este tipo, descrita por ejemplo en el capítulo 16 del manual "Automatic Speech and Speaker Recognition" editado por Kluwer Academic Publishers, permite reducir considerablemente, con relación al modelo utilizado en el marco de la primera aproximación descrita más arriba, los tiempos individuales de los procesos de aprendizaje del modelo sublexical y del modelo de articulación, pues cada uno de estos modelos presenta una estructura simple con relación al modelo lexical utilizado en la primera aproximación.

Los modos de realización conocidos de esta segunda aproximación recurren lo más a menudo a un primer y a un segundo transductor, cada uno formado por un modelo de Markov representativo de una cierta fuente de conocimientos, es decir, para retomar el caso mencionado anteriormente, un primer modelo de Markov representativo de las entidades

sublexicales y un segundo modelo de Markov representativo de combinaciones posibles de las indicadas entidades sublexicales. En el transcurso de una etapa de decodificado de datos de entrada, se activarán estados contenidos en los primero y segundo transductores, cuyos estados son respectivamente representativos de modelizaciones posibles de las entidades sublexicales a identificar y de modelizaciones posibles de combinaciones de las indicadas entidades sublexicales. Los estados activados de los primero y segundo transductores se memorizarán entonces en medios de memorización.

Según una representación conceptual elegante de esta segunda aproximación, los primero y segundo transductores pueden estar representados en forma de un transductor único equivalente a los primero y segundo transductores tomados en su composición, permitiendo traducir los datos de entrada en entidades lexicales, explotando simultáneamente el modelo sublexical y el modelo de articulación.

Según esta representación conceptual, la memorización de los estados activados en el transcurso de la etapa de decodificado equivale a una memorización de estados de este transductor único, del cual cada estado puede ser considerado por una parte, como un par formado por un estado del primer transductor formado por el primer modelo construido en base a entidades sublexicales, y por otra parte por un estado del segundo transductor formado por el segundo modelo construido en base a entidades lexicales. Una memorización de este tipo podría ser realizada de forma anárquica, a medida que estos se fueran produciendo. Sin embargo, el número máximo de estos diferentes que puede tomar el transductor único es muy grande, pues es igual a un producto entre los números máximos de estados que pueden tomar cada uno de los primero y segundo transductores. Por otro lado, el número de estados del transductor único efectivamente útiles para el decodificado, es decir que corresponden efectivamente a secuencias sublexicales y lexicales permitidas en la lengua considerada, es relativamente bajo con relación al número máximo de estados posibles, particularmente si estados cuya activación es poco probable, aunque teóricamente permitida, son excluidos por convención. Así, una memorización anárquica de los estados producidos por el transductor único conduce a utilizar una memoria de tamaño importante, en la cual las informaciones representativas de los estados producidos estarían muy diseminadas, lo cual conduciría a utilizar para su diseccionado para fines de lectura y/o de escritura números de gran tamaño que necesitan un sistema de gestión de acceso de memoria indebidamente complejo con relación al volumen de informaciones útiles efectivamente contenido en la memoria, que inducirá tiempos de acceso de memoria importantes e incompatibles con tensiones temporales propias por ejemplo para aplicaciones de traducción en tiempo real.

La invención, tal como se ha reivindicado en las reivindicaciones 1 y 8, tiene por objeto remediar en una amplia medida este inconveniente, proponiendo un procedimiento de traducción de datos que utilice un transductor único y medios de memorización destinados a contener informaciones relativas a los estados activados de dicho transductor único, procedimiento gracias al cual accesos de lectura/escritura a las indicadas informaciones pueden ser realizados lo suficientemente rápido como para permitir una utilización de dicho procedimiento en aplicaciones de traducción en tiempo real.

En efecto, según la invención, un procedimiento de traducción de datos de entrada en al menos una secuencia lexical de salida incluye una etapa de decodificado de los datos de entrada en el transcurso de la cual entidades sublexicales de las cuales los mencionados datos son representativas se identifican por medio de un primer modelo construido en base a entidades sublexicales predeterminadas, y en el transcurso de la cual se generan, a medida que se van identificando las entidades sublexicales y en referencia a al menos un segundo modelo construido en base a entidades lexicales, diversas combinaciones posibles de las mencionadas entidades sublexicales, estando cada combinación destinada para ser memorizada, conjuntamente con un valor de verosimilitud asociado, en medios de memorización que incluyen una pluralidad de zonas de memoria de las cuales cada una es apta para contener al menos una de las indicadas combinaciones, estando cada zona provista de una dirección igual a un valor tomado por una función escalar predeterminada cuando la indicada función se aplica a parámetros adecuados a entidades sublexicales y a su combinación destinadas para ser memorizadas juntas en la zona considerada.

La utilización de zonas de memoria direccionadas por medio de una función escalar predeterminada permite organizar el almacenado de las informaciones útiles producidas por este transductor único y simplificar la gestión de los accesos a estas informaciones ya que, conforme a la invención, la memoria se subdivide en zonas destinadas cada una a contener informaciones relativas a estados efectivamente producidos por el transductor único. Esto permite un direccionado de las indicadas zonas por medio de un número cuyo tamaño es reducido con relación al tamaño necesario para el diseccionado de una memoria concebida para memorizar de forma anárquica cualquier par de estados de los primero y segundo transductores.

En un modo de realización ventajoso de la invención, se eligirá por función escalar predeterminada una función esencialmente inyectiva, es decir una función, que, aplicada a diferentes parámetros tomará salvo excepciones valores diferentes, lo cual permite asegurar que cada zona de memoria solo contendrá en principio informaciones relacionadas como máximo a una sola combinación de entidades sublexicales, es decir a un solo estado del transductor equivalente, lo cual permite simplificar también los accesos a las indicadas informaciones suprimiendo la necesidad de una selección, dentro de una misma zona de memoria, entre informaciones relativas a diferentes combinaciones de entidades sublexicales.

En una variante de este modo de realización, la función escalar predeterminada será además igualmente esencialmente sobreyectiva además de ser inyectiva, es decir que cada zona de memoria disponible está destinada para contener efectivamente, salvo excepción, informaciones relacionadas con una sola combinación de entidades sublexi-

cales, lo cual representa una utilización óptima de los medios de memorización ya que su potencial de memorización se explotará entonces plenamente. En esta variante, la función escalar predeterminada será de hecho esencialmente biyectiva, en tanto que a la vez esencialmente inyectiva y sobreyectiva.

Los parámetros de entradas de la función escalar predeterminada pueden adoptar múltiples formas según el modo de realización de la invención elegido. En uno de estos modos de realización, el modelo sublexical contiene modelos de entidades sublexicales de los cuales diferentes estados se numeran de forma contigüa y presentan un número total inferior o igual a un primer número predeterminado adecuado al modelo sublexical, y el modelo de articulación contiene modelos de combinaciones posibles de entidades sublexicales cuyos diferentes estados se numeran de forma contigua y presentan un número total inferior o igual a un segundo número predeterminado adecuado al modo de articulación, constituyendo los números de los estados de las entidades sublexicales y de sus combinaciones posibles los parámetros a los cuales la función escalar predeterminada está destinada para ser aplicada.

La función escalar predeterminada puede adoptar múltiples formas según el modo de realización de la invención elegido. En un modo de realización particular de la invención, cada valor tomado por la función escalar predeterminada es una concatenación de un resto de una primera división entera por el primer número predeterminado del número de un estado de una entidad sublexical identificado por medio del primer modelo y de un resto de una segunda división entera por el segundo número predeterminado del número de un estado de una combinación identificado por medio del segundo modelo.

Una concatenación de este tipo garantiza en principio que los valores de los restos de las primera y segunda divisiones enteras se utilizaran sin alteración para los fines de direccionado de las zonas de memoria, produciendo así una reducción máxima de un riesgo de error en el direccionado.

En un modo de realización particularmente ventajoso de la invención, en cuanto a que utiliza medios probados e individualmente conocidos por el experto en la materia, la etapa de decodificado utiliza un algoritmo de Viterbi aplicado conjuntamente a un primer modelo de Markov que presenta estados representativos de diferentes modelizaciones posibles de cada entidad sublexical permitida en una lengua de traducción dada, y en un segundo modelo de Markov que presenta estados representativos de diferentes modelizaciones posibles de cada articulación entre dos entidades sublexicales permitida en la indicada lengua de traducción.

En un aspecto general, la invención se refiere igualmente a un procedimiento de traducción de datos de entrada en una secuencia lexical de salida, que incluye una etapa de decodificado de los datos de entrada destinada para ser realizada por medio de un algoritmo del tipo algoritmo de Viterbi, explotando simultáneamente una pluralidad de fuentes de conocimientos distintas que forman un transductor único cuyos estados están destinados para ser memorizados, conjuntamente con un valor de verosimilitud asociado, en medios de memorización que incluyen una pluralidad de zonas de memoria de las cuales cada una es apta para contener al menos uno de los indicados estados, estando cada zona provista de una dirección igual a un valor tomado por una función escalar predeterminada cuando la indicada función se aplica a parámetros adecuados a los estados de dicho transductor único.

La invención se refiere igualmente a un sistema de reconocimiento de señales acústicas que utiliza un procedimiento tal como el descrito anteriormente.

Las características de la invención mencionadas anteriormente, así como otras, aparecerán más claramente con la lectura de la descripción siguiente de un ejemplo de realización, realizándose la indicada descripción en relación con los dibujos adjuntos, entre los cuales:

La figura 1 es un esquema conceptual que describe un decodificador en el cual se realiza un procedimiento conforme a la invención.

La figura 2 es un esquema que describe la organización de una tabla destinada a memorizar informaciones producidas por dicho decodificador.

La figura 3 es un esquema funcional que describe un sistema de reconocimiento acústico conforme a un modo de realización particular de la invención.

La figura 4 es un esquema funcional que describe un primer decodificador destinado para realizar dentro de este sistema una primera etapa de decodificado, y

La figura 5 es un esquema funcional que describe un segundo decodificador destinado para realizar dentro de este sistema una segunda etapa de decodificado conforme al procedimiento según la invención.

La figura 1 representa un decodificador DEC destinado a recibir datos de entrada AVin y para proporcionar una secuencia lexical de salida LSQ. Este decodificador DEC incluye una máquina de Viterbi VM, destinada para realizar un algoritmo de Viterbi conocido del experto en la materia, cuya máquina de Viterbi VM utiliza conjuntamente un primer modelo de Markov APHM representativo de todas las modelizaciones posibles de cada entidad sublexical permitida en una lengua de traducción dada, y un segundo modelo de Markov PHLM representativo de todas las modelizaciones posibles de cada articulación entre dos entidades sublexicales permitida en la indicada lengua de traducción, cuyos pri-

mero y segundo modelos de Markov APHM y PHLM pueden respectivamente ser representados en forma de un primer transductor T1 destinado a convertir secuencias de vectores acústicos en secuencias de entidades sublexicales Phsq, por ejemplo fonemas, y en forma de un segundo transductor T2 destinado para convertir estas secuencias de entidades sublexicales Phsq en secuencias lexicales LSQ, es decir en este ejemplo en secuencias de palabras. Cada transductor T1 o T2 puede ser asimilado a un autómata enriquecido con estados acabados, correspondiendo cada estado ei o ej respectivamente a un estado de una entidad sublexical o a un estado de una combinación de tales entidades identificadas por el primero o segundo transductor T1 ó T2. En una representación conceptual de este tipo, el decodificador DEC es por consiguiente un transductor único, equivalente a una composición de los primero y segundo transductores T1 y T2, que explota simultáneamente el modelo sublexical y el modelo de articulación y produce estados (ei; ej) de los cuales cada uno es un par formado por un estado ei del primer transductor T1, por una parte, y por un estado ej del segundo transductor T2, por otra parte, siendo un estado (ei; ej) por si mismo representativo de una combinación posible de entidades sublexicales. Conforme a la invención, cada estado (ei; ej) está destinado para ser memorizado, conjuntamente con un valor de verosimilitud Sij asociado, en medios de memorización, constituidos en este ejemplo por una tabla TAB.

La figura 2 representa esquemáticamente una tabla TAB, que incluye una pluralidad de zonas de memoria MZ1, MZ2, MZ3...MZN, de las cuales cada una es apta para contener al menos uno de los indicados estados (e1i; e2j) del transductor único, acompañado del valor de verosimilitud Sij que le ha sido atribuido. Cada zona MZ1, MZ2, MZ3...MZN está provista de una dirección igual a un valor tomado por una función escalar h predeterminada cuando la indicada función se aplica a parámetros adecuados a entidades sublexicales y a su combinación destinada para ser memorizada en la zona considerada.

En el modo de realización de la invención descrito aquí, la función escalar h es una función esencialmente inyectiva, es decir una función que, aplicada a diferentes parámetros tomará salvo excepción valores diferentes, lo cual permite asegurar que cada zona de memoria MZm (para m = 1 a N) solo contendrá en principio informaciones relacionadas como máximo a una sola combinación de entidades sublexicales, es decir a un solo estado (ei; ej) del transductor formado por el decodificador descrito anteriormente. La función escalar h es además igualmente esencialmente sobreyectiva en este ejemplo, es decir que cada zona de memoria MZm (para m = 1 a N) está destinada a contener efectivamente, salvo excepción, informaciones relacionadas con un estado (ei; ej) de dicho transductor. La función escalar h es por consiguiente aquí esencialmente biyectiva, mientras que a la vez esencialmente inyectiva y esencialmente sobreyectiva. Cuando el transductor produce un nuevo estado (ex; ey), bastará, para saber si esta composición de estados de los primero y segundo transductores ha sido ya producida, y con que verosimilitud, interrogar la tabla TAB por medio de la dirección h[(ex; ey)]. Si esta dirección corresponde a una zona de memoria MZm ya definida en la tabla para un estado (ei; ej), se establecerá una identidad entre el nuevo estado (ex; ey) y el estado (ei; ej) ya memorizado.

En este modo de realización, el modelo sublexical contiene diferentes modelizaciones posibles ei de cada entidad sublexical, numeradas de forma contigua y presentando un número total inferior o igual a un primer número predeterminado V1 adecuado al modelo sublexical, y el modelo de articulación contiene diferentes modelizaciones posibles ej de posibles combinaciones de estas entidades sublexicales, numeradas de forma contigua y presentando un número total inferior o igual a un segundo número predeterminado V2 adecuado al modelo de articulación, constituyendo los números de las entidades sublexicales y de sus combinaciones posibles los parámetros a los cuales la función escalar h predeterminada está destinada para ser aplicada.

Cada valor tomado por la función escalar predeterminado es una concatenación de un resto, que puede variar de 0 a (V1-1), de una primera división entera por el primer número predeterminado V1 del número de la modelización de un estado de una entidad sublexical identificado por medio del primer modelo y de un resto, que puede variar de 0 a (V2-1), de una segunda división entera por el segundo número predeterminado V2 del número de la modelización de un estado de una combinación de entidades sublexicales identificado por medio del segundo modelo. Así, si en un ejemplo irrealista por consiguiente simplificado en extremo para permitir una comprensión fácil de la invención, las entidades sublexicales modelizadas en el primer modelo de Markov son tres fonemas “p”, “a” y “o”, de los cuales cada uno puede ser modelizado por cinco estados distintos, es decir estados (ei = 0, 1, 2, 3 ó 4) para el fonema “p”, estados (ei = 5, 6, 7, 8 ó 9) para el fonema “a”, y estados (ei = 10, 11, 12, 13 ó 14) para el fonema “o”, el primer número predeterminado V1 será igual a 5.

Si las combinaciones de entidades sublexicales modelizadas en el segundo modelo de Markov son dos combinaciones “pa” y “po”, de las cuales cada una puede modelizarse por dos estados distintos, es decir estados (ej = 0 ó 1) para la combinación “pa”, y estados (ej = 2 ó 3) para la combinación “po”, el segundo número predeterminado será igual a 4.

Las diferentes modelizaciones posibles en entidades sublexicales y de sus combinaciones se encuentran como máximo en número de N = 20, la dirección h[(0; 0)] de la primera zona de memoria MZ1 tendrá por valor la concatenación del resto de la división entera 0/V1 = 0 con el resto de la división entera 0/V2 = 0 o la concatenación 00 de un valor 0 con un valor 0. La dirección h[(14; 3)] de la Nª zona de memoria MZN tendrá por valor la concatenación del resto de la división entera de 14 por V1 (con V1 = 5) con el resto de la división entera de 3 por V2 (con V2 = 4), o sea la concatenación 43 de un valor 4 con un valor 3.

Una concatenación de este tipo garantiza en principio que los valores de los restos de las primera y segunda divisiones enteras se utilizarán sin alteración para los fines de direccionado de las zonas de memoria, produciendo así una reducción máxima de un riesgo de error en el direccionado. Sin embargo, una concatenación de este tipo conduce a utilizar números hechos artificialmente más grandes de lo necesario con relación al número de zonas de memoria N efectivamente direccionadas. Técnicas, conocidas del experto en la materia, permiten comprimir números a concatenar limitando las pérdidas de información relacionadas con dicha compresión. Se podrá por ejemplo prever hacer que se monten representaciones binarias de los indicados números, realizando una operación OU-EXCLUSIF entre bitios de peso bajo de uno de estos números binarios con los bitios de peso fuerte del otro número binario.

Con el fin de facilitar su comprensión, la descripción de la invención que antecede ha sido realizada de un ejemplo de aplicación donde una máquina de Viterbi opera sobre un transductor único formado por una composición de dos modelos de Markov. Esta descripción se puede generalizar en aplicaciones donde una única máquina de Viterbi explota simultáneamente un número P superior a 2 de fuentes de conocimientos diferentes, formando así un transductor destinado para producir estados (e1i; e2j;...; ePs), pudiendo cada uno de los cuales memorizarse en una zona de memoria de una tabla, cuya zona de memoria se identificará por medio de una dirección h[(e1i; e2j;...; ePs)] o h es una función escalar predeterminada tal como se ha descrito más arriba.

La figura 3 representa esquemáticamente un sistema SYST de reconocimiento acústico según un modo de realización particular de la invención, destinado a traducir una señal acústica de entrada ASin en una secuencia lexical de salida OUTSQ. En este ejemplo, la señal de entrada ASin está constituida por una señal electrónica analógica, que podrá provenir por ejemplo de un micrófono no representado en la figura. En el modo de realización descrito aquí, el sistema SYST incluye una etapa de entrada FE, que contiene un dispositivo de conversión analógico/digital ADC, destinado a proporcionar una señal digital ASin(1:n), formada por muestras ASin(1), ASin(2)... ASin(n) codificadas cada una sobre b bitios, y representativa de la señal acústica de entrada ASin, y un módulo de muestreo SA, destinado a convertir la señal acústica digital ASin(1:n) en una secuencia de vectores acústicos AVin, estando cada vector provisto de componentes AV1, AV2...AVr donde r es la dimensión de un espacio acústico definido para una aplicación dada a la cual el sistema de traducción SYST está destinado, evaluándose cada una de las componentes AVi (para i=1 a r) en función de características adecuadas a este espacio acústico. En otros modos de realización de la invención, la señal de entrada ASin podrá, desde el origen, ser de naturaleza digital, lo cual permitirá eludir la presencia del dispositivo de conversión analógica/digital ADC dentro de la etapa de entrada FE.

El sistema SYST incluye además un primer decodificador DEC1, destinado a proporcionar una selección Int1, Int2... IntK de interpretaciones posibles de la secuencia de vectores acústicos AVin en referencia a un modelo APHM construido en base a entidades sublexicales predeterminadas.

El sistema SYST incluye además un segundo decodificador DEC2 en el cual un procedimiento de traducción conforme a la invención es realizado con miras a analizar datos de entrada constituidos por los vectores acústicos. AVin en referencia a un primer modelo construido en base a entidades sublexicales predeterminadas, por ejemplo extraído del modelo APHM, y en referencia a un segundo modelo construido en base a modelizaciones acústicas procedentes de una biblioteca BIB. El segundo decodificador DEC2 identificará así la de las mencionadas interpretaciones Int1, Int2... IntK que deberá constituir la secuencia lexical de salida OUTSQ.

La figura 4 representa más en detalle el primer decodificador DEC1, que incluye una primera máquina de Viterbi VM1, destinada a ejecutar una primera subetapa de decodificado de la secuencia de vectores acústicos AVin representativa de la señal acústica de entrada y previamente generada por la etapa de entrada FE, cuya secuencia se memorizará además ventajosamente en una unidad de almacenado MEM1 por razones que aparecerán en lo que sigue de la exposición. La primera subetapa de decodificado se realiza en referencia a un modelo de Markov APMM que permite en bucle todas las entidades sublexicales, de preferencia todos los fonemas de la lengua en la cual la señal acústica de entrada debe traducirse si se considera que las entidades lexicales son palabras, estando las entidades sublexicales representadas en forma de vectores acústicos predeterminados.

La primera máquina de Viterbi VM1 es apta para restituir una secuencia de fonemas Phsq que constituye la traducción fonética más próxima de la secuencia de vectores acústicos AVin. Los tratamientos ulteriores realizados por el primer decodificador DEC1 se realizarán así a nivel fonético, y ya no a nivel vectorial, lo cual reduce considerablemente la complejidad de los indicados tratamientos, siendo cada vector una entidad multidimensional que presenta r componentes, mientras que un fonema puede en principio ser identificado por una etiqueta unidimensional que le es adecuada, como por ejemplo una etiqueta “OU” atribuida a una vocal oral “u”, o una etiqueta “CH” atribuida a una consonante frictiva no-sonora “$\int$”. La secuencia de fonemas Phsq generada por la primera máquina de Viterbi VM1 está así constituida por una sucesión de etiquetas más fácilmente manipulables que no lo serían con vectores acústicos.

El primer decodificador DEC1 incluye una segunda máquina de Viterbi VM2 destinada para realizar una segunda subetapa de decodificado de la secuencia de fonemas Phsq generada por la primera máquina de Viterbi VM1. Esta segunda etapa de decodificado se realiza con referencia a un modelo de Markov PLMM constituido por transcripciones sublexicales de entidades lexicales, es decir en este ejemplo de transcripciones fonéticas de palabras presentes en el vocabulario de la lengua en el cual la señal acústica de entrada debe traducirse. La segunda máquina de Viterbi está destinada para interpretar la secuencia de fonemas Phsq, que es fuertemente ruidosa debido a que el modelo APMM utilizado por la primera máquina Viterbi VM1 es de una gran sencillez, y utiliza predicciones y comparaciones entre

secuencias de etiquetas de fonemas contenidos en la secuencia de fonemas Phsq y diversas combinaciones posibles de etiquetas de fonemas previstas en el modelo de Markov PLMM. Aunque una máquina de Viterbi restituya solo usualmente la de las secuencias que presentan la mayor probabilidad, la segunda máquina de Viterbi VM2 utilizada aquí restituirá ventajosamente todas las secuencias de fonemas Isq1, Isq2... IsqN que la indicada segunda máquina VM2 habrá podido reconstituir, con valores de probabilidad asociados p1, p2... pN que habrán sido calculados para las indicadas secuencias y serán representativas de la fiabilidad de las interpretaciones de la señal acústica que estas secuencias representan.

Todas las interpretaciones posibles Isq1, Isq2... IsqN al volverse automáticamente disponibles al inicio de la segunda subetapa de decodificado, una selección realizada por un módulo de selección SM de las K interpretaciones Int1, Int2... IntK que presentan los valores más fuertes de probabilidad es simplificada sea cual fuere el valor de K que hubiera sido elegido.

Los modelos de Markov APMM y PLMM pueden ser considerados como sub-conjuntos del modelo APHM mencionado más arriba.

Las primera y segunda máquinas de Viterbi VM1 y VM2 pueden funcionar en paralelo, generando entonces la primera máquina de Viterbi VM1 a medida que se van produciendo etiquetas de fonemas que serán inmediatamente tomados en cuenta por la segunda máquina de Viterbi VM2, lo cual permite reducir el tiempo total percibido por un usuario del sistema necesario para la combinación de las primera y segunda subetapas de decodificado permitiendo la realización del conjunto de recursos de cálculo necesarios para el funcionamiento del primer decodificador DEC1 una vez que los vectores acústicos AVin representativos de la señal acústica de entrada aparezcan, y no después de que hayan sido completamente traducidos en una secuencia completa de fonemas Phsq por la primera máquina de Viterbi VM1.

La figura 5 representa más en detalle un segundo decodificador DEC2 conforme a un modo de realización particular de la invención. Este segundo decodificador DEC2 incluye una tercera máquina de Viterbi VM3 destinada para analizar la secuencia de vectores acústicos AVin representativa de la señal acústica de entrada que ha sido previamente memorizada a este efecto en la unidad de almacenado MEM1.

A este respecto, la tercera máquina de Viterbi VM3 está destinada para identificar las entidades sublexicales cuyos vectores acústicos A Vin son representativos por medio de un primer modelo construido en base a entidades sublexicales predeterminadas, en este ejemplo el modelo de Markov APMM realizado en el primer decodificador y ya descrito más arriba, y para producir estados eli representativos de las entidades sublexicales así identificadas. Una explotación del modelo de Markov APMM de este tipo puede ser representada como una realización de un primer transductor T1 parecido al descrito más arriba.

La tercera máquina de Viterbi VM3 genera además, a medida que se identifican entidades sublexicales y en referencia a al menos un modelo de Markov específico PHLM construido sobre la base de entidades lexicales, diversas combinaciones posibles de las entidades sublexicales, y en producir estados e2j representativos de las combinaciones de entidades sublexicales así generadas, estando la combinación más verosímil destinada para formar la secuencia lexical de salida OUTSQ. Una explotación de este tipo del modelo de Markov PHLM puede ser representada como una realización de un segundo transductor T2 similar al descrito más arriba.

La explotación simultánea de los modelos de Markov APMM y PHLM por la tercera máquina de Viterbi VM3 puede por consiguiente ser tomada como la utilización de un transductor único formado por una composición de los primero y segundo transductores tales como los descritos más arriba, destinado a producir estados (ei; ej) provistos cada uno de un valor de verosimilitud Sij. Conforme a la descripción de la invención que antecede, estos estados se memorizaran en una tabla TAB incluida en una unidad de almacenado MEM2, que podrá formar parte de una memoria central o de una memoria oculta que incluye igualmente la unidad de almacenado MEM1, siendo cada estado (ei; ej) almacenado con su valor de verosimilitud asociada Sij en una zona de memoria que tiene por dirección un valor h[(ei; ej)], con las ventajas en términos de rapidez de acceso anteriormente mencionadas. Un decodificador de memoria MDEC seleccionará al inicio del proceso de decodificado el de las combinaciones de entidades sublexicales memorizadas en la tabla TAB que presentará la mayor verosimilitud, es decir el mayor valor de Sij, destinado a formar la secuencia lexical de salida OUTSQ.

El modelo de Markov específico PHLM está aquí especialmente generado por un módulo de creación de modelo MGEN, y es únicamente representativo de ensamblajes posibles de fonemas dentro de las secuencias de palabras formadas por las diversas interpretaciones fonéticas Int1, Int2,... IntK de la señal acústica de entrada proporcionadas por el primer decodificador, cuyos ensamblajes están representados por modelizaciones acústicas que provienen de una biblioteca BIB de las entidades lexicales que corresponden a estas interpretaciones. El modelo de Markov específico PHLM presenta por consiguiente un tamaño limitado debido a su especificidad.

De este modo, los accesos a las unidades de almacenado MEM1 y MEM2, así como a los diferentes modelos de Markov utilizados en el ejemplo de realización de la invención descrito anteriormente necesitan una gestión poco compleja, debido a la sencillez de estructura de los indicados modelos y del sistema de direccionado de las informaciones destinadas a ser memorizadas y leídas en las indicadas unidades de almacenado. Estos accesos de memoria pueden por

consiguiente ser ejecutados lo suficientemente rápido para hacer el sistema descrito en este ejemplo apto para realizar traducciones en tiempo real de datos de entrada en secuencias lexicales de salida.

Aunque la invención haya sido descrita aquí en el marco de una aplicación dentro de un sistema que incluye dos decodificadores dispuestos en cascada, es completamente considerable, en otros modos de realización de la invención, utilizar un único decodificador similar al segundo decodificador descrito más arriba, que podrá por ejemplo realizar un análisis acústico-fonético y memorizar, a medida que se van identificando fonemas, diversas combinaciones posibles de los indicados fonemas, estando la combinación de fonemas más verosímil destinada a formar la secuencia lexical de salida.

## REIVINDICACIONES

1. Procedimiento de traducción de datos de voz en al menos una secuencia lexical de salida, que incluye una etapa de decodificado de los datos de voz en el transcurso de la cual entidades sublexicales de las cuales los indicados datos son representativas se identifican por medio de un primer modelo construido en base a entidades sublexicales predeterminadas, y en el transcurso de la cual son generadas, a medida que se van identificando las entidades sublexicales y en referencia a al menos un segundo modelo construido en base a entidades lexicales, diversas combinaciones posibles de las indicadas entidades sublexicales, estando cada combinación destinada para ser memorizada, conjuntamente con un valor de verosimilitud asociado, en medios de memorización que incluyen una pluralidad de zonas de memoria de las cuales cada una es apta para contener al menos una de las indicadas combinaciones, estando cada zona provista de una dirección igual a un valor tomado por una función escalar predeterminada cuando la indicada función se aplica a parámetros adecuados a entidades sublexicales y a su combinación destinadas para ser memorizadas juntas en la zona considerada.

2. Procedimiento de traducción según la reivindicación 1, en el cual la función escalar predeterminada es una función de naturaleza inyectiva.

3. Procedimiento de traducción según la reivindicación 2, en el cual la función escalar predeterminada es además igualmente de naturaleza sobreyectiva.

4. Procedimiento de traducción según la reivindicación 1, en el cual el modelo sublexical contiene modelos de entidades sublexicales de las cuales diferentes estados son numerados de forma contigua y presentan un número total inferior o igual a un primer número predeterminado adecuado al modelo sublexical, y en el cual el modelo de articulación contiene modelos de combinaciones posibles de entidades sublexicales cuyos diferentes estados son numerados de forma contigua y presentan un número total inferior o igual a un segundo número predeterminado adecuado al modelo de articulación, constituyendo los números de estados de las entidades sublexicales y sus combinaciones posibles los parámetros a los cuales la función escalar predeterminada está destinada para ser aplicada.

5. Procedimiento de traducción según la reivindicación 4, en el cual cada valor tomado por la función escalar predeterminada es una concatenación de un resto de una primera división entera por el primer número predeterminado del número de un estado de una entidad sublexical identificado por medio del primer modelo y de un resto de una segunda división entera por el según número predeterminado del número de un estado de una combinación identificado por medio del segundo modelo.

6. Procedimiento de traducción según una de las reivindicaciones 1 a 5, según el cual la etapa de decodificado utiliza un algoritmo de Viterbi aplicado conjuntamente a un primer modelo de Markov que presenta estados representativos de diferentes modelizaciones posibles de cada entidad sublexical autorizada en una lengua de traducción dada, y a un segundo modelo de Markov que presenta estados representativos de diferentes modelizaciones posibles de cada articulación entre dos entidades sublexicales autorizadas en la indicada lengua de traducción.

7. Procedimiento de traducción de datos de voz en una secuencia lexical de salida, que incluye una etapa de decodificado de los datos de voz destinada para ser ejecutada por medio de un algoritmo del tipo algoritmo de Viterbi, que explota simultáneamente una pluralidad de fuentes de conocimientos distintas que forman un transductor único cuyos estados están destinados para ser memorizados, conjuntamente con un valor de verosimilitud asociado, en medios de memorización que incluyen una pluralidad de zonas de memoria de las cuales cada una es apta para contener al menos uno de los indicados estados, estando cada zona provista de una dirección igual a un valor tomado por una función escalar predeterminada cuando la indicada función se aplica a parámetros adecuados a los estados de dicho transductor único.

8. Sistema de reconocimiento de vocal que comprende medios para realizar un procedimiento de traducción conforme a una de las reivindicaciones 1 a 7.

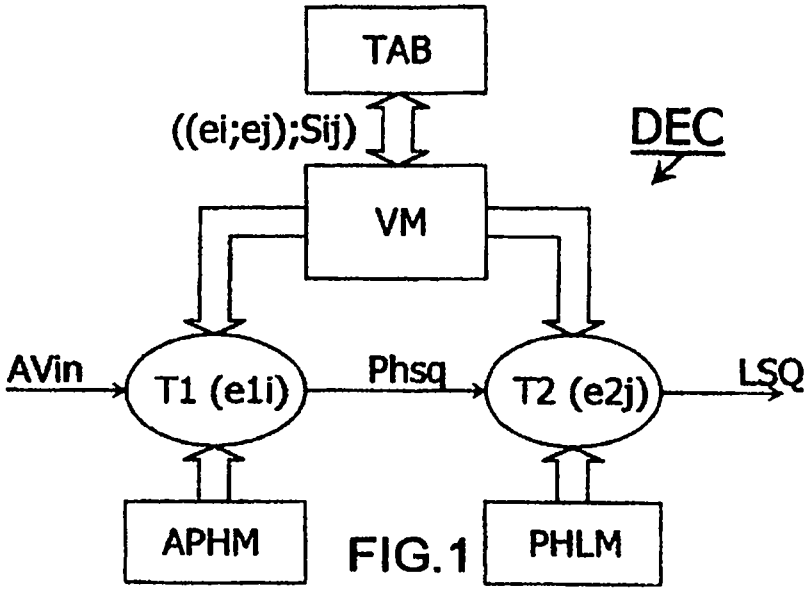


FIG.1

TAB

| h[(ei;ej)] | ((ei;ej);Sij) | |
|---|---|---|
| 00 | ((0;0);S00) | MZ1 |
| 10 | ((1;0);S10) | MZ2 |
| 20 | ((2;0);S20) | MZ3 |
| ⋮ | ⋮ | |
| 43 | ((4;3);S43) | MZN |

FIG.2

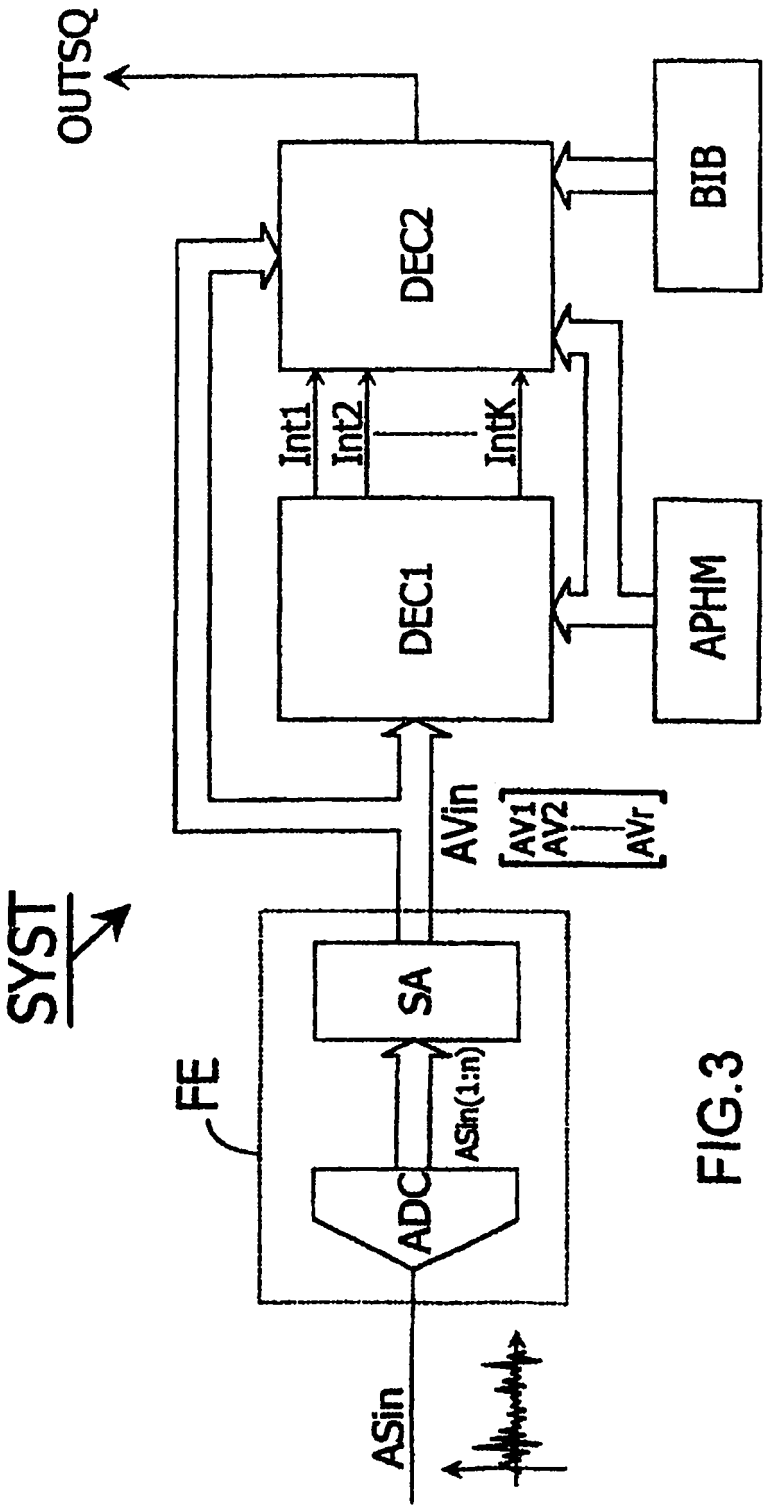
OUTSQ
BIB
DEC2
Int1
Int2
IntK
DEC1
APHM
AVin
AV1
AV2
AVr
SYST
SA
FE
ASIn(1:n)
ADC
ASin

FIG.3

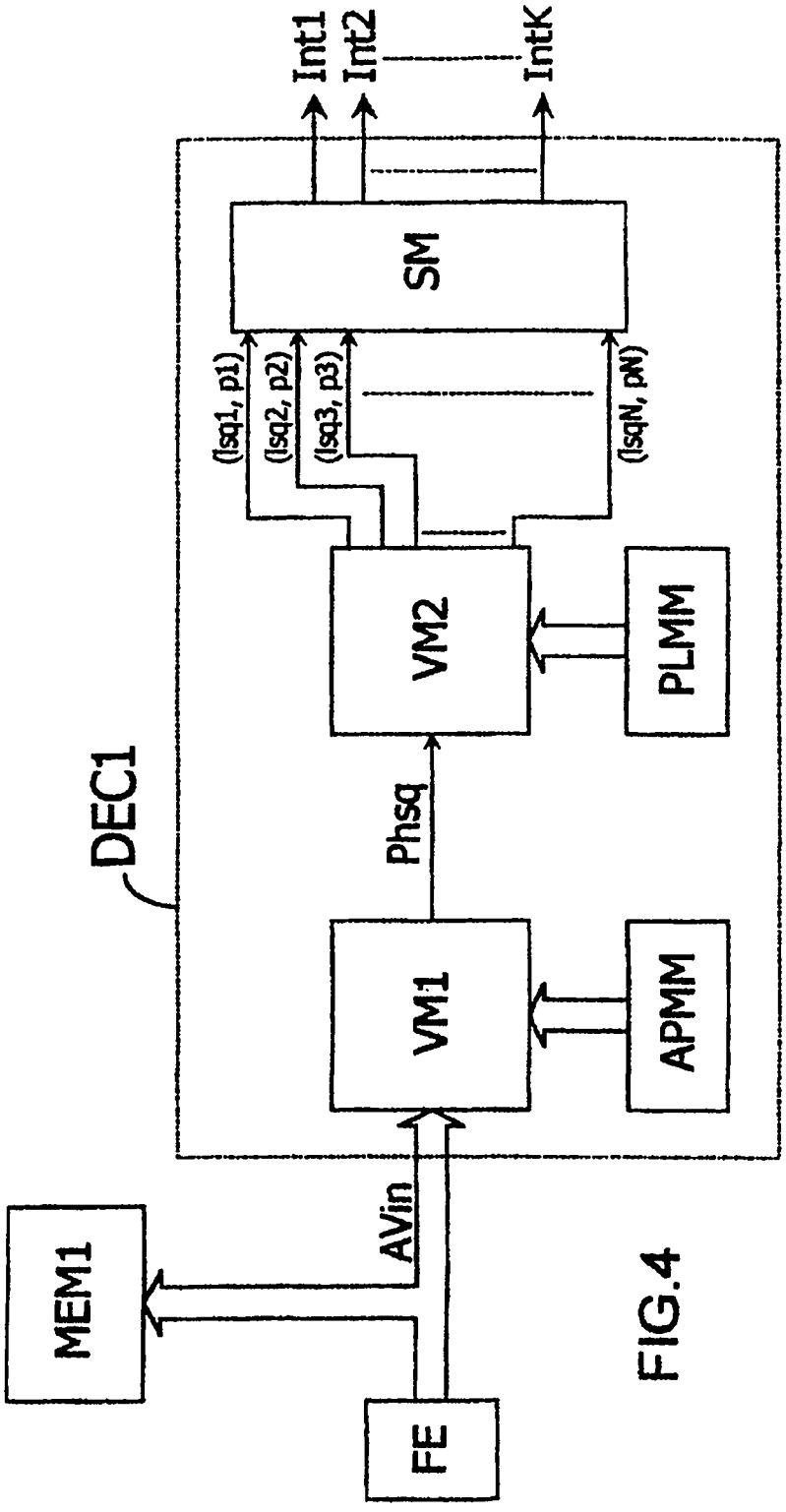
Int1
Int2
IntK
SM
(Isq1, p1)
(Isq2, p2)
(Isq3, p3)
(IsqN, pN)
VM2
PLMM
DEC1
Phsq
VM1
APMM
AVin
MEM1
FE

FIG.4

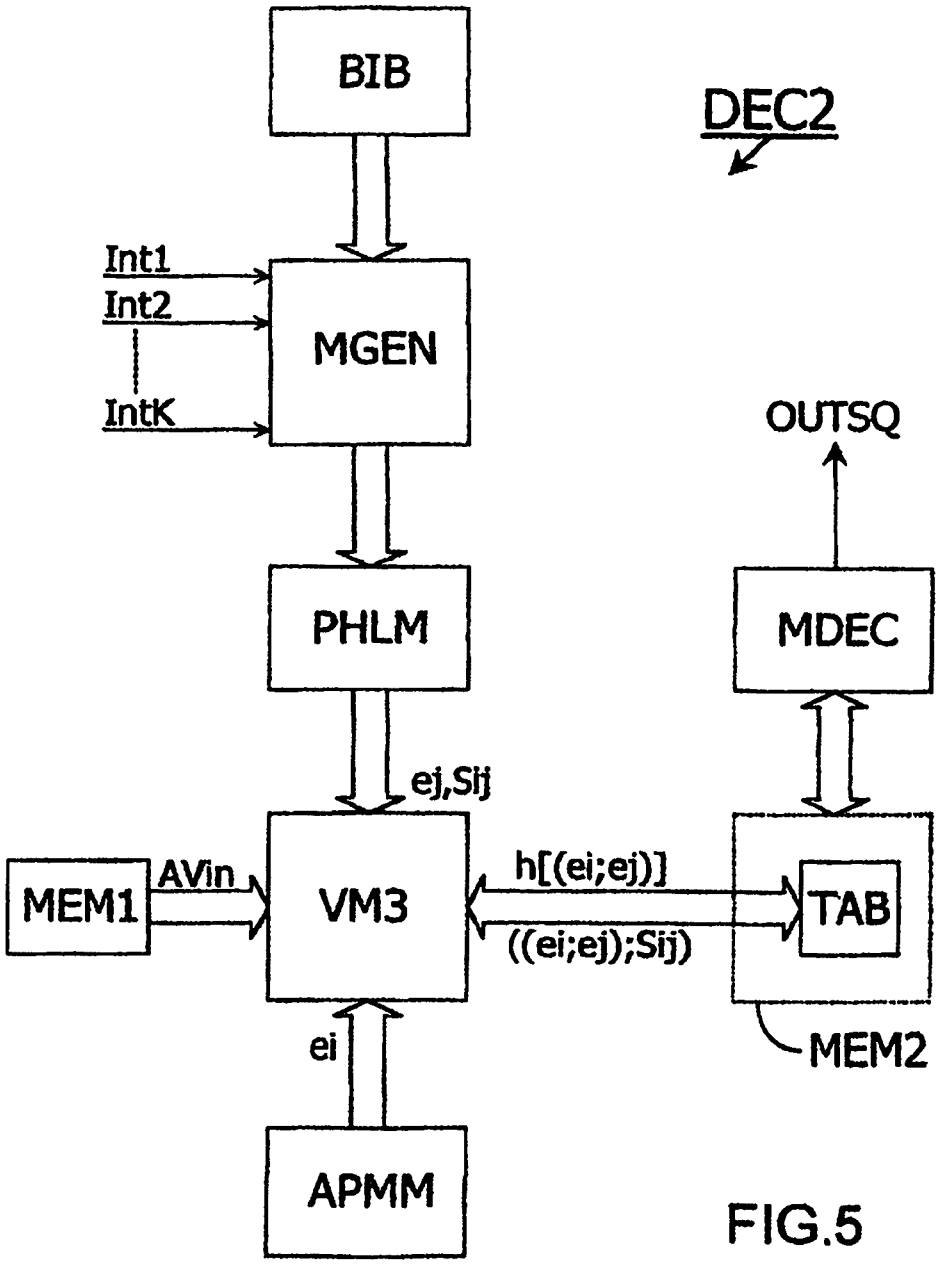
BIB
DEC2
Int1
Int2
IntK
MGEN
OUTSQ
PHLM
MDEC
ej,Sij
MEM1
AVin
VM3
h[(ei;ej)]
((ei;ej);Sij)
TAB
MEM2
ei
APMM

FIG.5