



República Federativa do Brasil
Ministério da Indústria, Comércio Exterior
e Serviços
Instituto Nacional da Propriedade Industrial

(11) PI 0304532-3 B1

(22) Data do Depósito: 15/04/2003

(45) Data de Concessão: 06/02/2018



(54) Título: "MÉTODO E APARELHO PARA SINCRONIZAÇÃO DE FLUXOS DE ÁUDIO E VÍDEO E MEIO LEGÍVEL POR COMPUTADOR"

(51) Int.Cl.: H04N 5/60

(30) Prioridade Unionista: 10/01/2003 US 10/340.477, 19/04/2002 US 60/374.269

(73) Titular(es): THOMSON LICENSING S.A.

(72) Inventor(es): CHARLES CHUANMING WANG; ANDREAS MATTHAIS AUST; SACHIN SATISH MODY

“MÉTODO E APARELHO PARA SINCRONIZAÇÃO DE FLUXOS DE ÁUDIO E VÍDEO e MEIO LEGÍVEL POR COMPUTADOR”

Referência Cruzada a Pedidos de Patente Relacionados

Este pedido de patente reivindica o benefício do
5 pedido de patente provisório U.S. nº de série 60/374269, depositado em 19 de abril de 2002, que está aqui incorporado na sua íntegra como referência.

Antecedentes da Invenção

1. Campo da Invenção

10 A invenção diz respeito ao campo de sistemas de comunicação multimídia e, mais especificamente, à minimização de erros de sincronização dos lábios induzidos por redes de transporte de atraso variável.

2. Antecedentes da Invenção

15 O problema de “sincronização dos lábios” ou sinc. dos lábios é bem conhecido. Resumidamente, erros temporais na apresentação de fluxos de áudio e vídeo por um dispositivo de apresentação podem resultar em uma condição pela qual informação de áudio é apresentada antes (com antecedência)
20 ou depois (com atraso) da informação de vídeo correspondente, resultando, por exemplo, em sincronização deficiente entre a representação de áudio da voz de um interlocutor e a representação de vídeo dos lábios do interlocutor.

As técnicas anteriores para solucionar o assim denominado problema de sincronização dos lábios são relativamente complexas e algumas vezes provocam degradação na
25 informação de áudio e/ou vídeo. Por exemplo, é de conhecimento baixar quadros de vídeo de maneira tal que seja induzido um

avanço temporal da geração da imagem de vídeo para corrigir assim um sinal de áudio com antecedência.

Erros de sincronização dos lábios podem ser provocados por muitas fontes. De particular interesse é o uso de redes de atraso variável, tais como a Internet e outras redes de comutação de pacotes. Em tais redes, informação de áudio e vídeo é transportada como fluxos separados e independentes. Durante o processamento do transporte antes da introdução desses fluxos na rede de atraso variável, um cabeçalho da camada de transporte que contém um carimbo temporal, bem como outros metadados (por exemplo, a taxa de amostragem do codificador, ordem de pacote e similares), é adicionado a alguns ou a todos os pacotes de transporte. Os carimbos temporais para informação de áudio e vídeo são tipicamente derivados de uma fonte comum, tal como um relógio de tempo real. Infelizmente, à medida em que os pacotes de áudio e vídeo percorrem uma rede de atraso variável, anormalidades temporais são comunicadas, pacotes são abaixados, a ordem dos pacotes não é preservada e o tempo de atraso do pacote varia por causa das condições de rede. O resultado disto tudo é erro de sincronização dos lábios dentro de fluxos de áudio e vídeo recebidos que passam através da rede de atraso variável.

Sumário da Invenção

A invenção compreende um método e aparelho para reduzir problemas de sincronização dos lábios em fluxos de áudio e vídeo correspondentes pela adaptação de uma distribuição estatística de erros temporais em uma faixa de erros



considerada prejudicial ou perceptível a um ouvinte.

Especificamente, um método de acordo com uma modalidade da invenção compreende: receber unidades de acesso de vídeo e unidades de acesso de áudio correspondentes, as unidades de acesso de vídeo e áudio correspondentes representando informação audiovisual que tende a apresentar um erro de sincronização dos lábios descrito por uma primeira função de distribuição de probabilidades (pdf); e atrasar temporariamente uma das unidades de acesso de áudio e vídeo recebido por unidade de acesso de quantidade de sincronismo, não atrasada e atrasada correspondente representando informação audiovisual que tende a apresentar erro de sincronização dos lábios descrito por uma segunda pdf, a segunda pdf utilizando uma parte maior de uma janela de tolerância de sincronização dos lábios do que a primeira pdf.

Em uma outra modalidade, um método para produzir fluxos de vídeo e áudio codificados adaptado para uso em uma rede de atraso variável compreende codificar informação de vídeo e áudio temporalmente correspondentes para produzir fluxos de vídeo e áudio codificados, cada um dos fluxos de vídeo e áudio codificados compreendendo uma pluralidade de respectivos pacotes de vídeo e áudio que inclui pacotes de vídeo e áudio com carimbo temporalizado e adaptar pelo menos um dos pacotes de vídeo com carimbo temporalizado e os pacotes de áudio com carimbo temporalizado por um fator de temporização para reduzir a probabilidade de um erro de sincronização dos lábios de vídeo com atraso.



Em uma outra modalidade, um estimador de pdf de erro de sincronização dos lábios é implementado em um receptor para estimar dinamicamente a pdf. Com base na pdf estimada, um tempo de atraso de áudio ideal é calculado em termos de função objeto. O atraso calculado é então introduzido no lado do receptor.

Descrição Resumida dos Desenhos

Os preceitos da presente invenção podem ser facilmente entendidos com consideração à descrição detalhada seguinte em conjunto com os desenhos anexos, em que:

A Figura 1 descreve um diagrama de blocos de nível superior de um sistema de comunicações;

a Figura 2 descreve um diagrama de blocos de nível superior de um controlador;

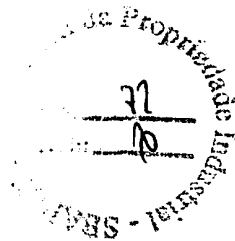
A Figura 3 descreve uma representação gráfica de uma função de densidade de probabilidades ($p(e)$) de um erro de sincronização dos lábios e útil no entendimento da presente invenção;

A Figura 4 descreve uma representação gráfica de uma janela de tolerância de erro de sincronização dos lábios (LSET) útil no entendimento da presente invenção;

A Figura 5 descreve uma representação gráfica de um deslocamento da pdf dentro de uma janela de tolerância;

A Figura 6 descreve um método para processar pacotes de áudio e/ou vídeo de acordo com a invenção;

A Figura 7 descreve um diagrama de blocos de nível superior de um sistema de comunicação de acordo com uma modalidade alternativa da invenção; e



A Figura 8 descreve um diagrama de blocos de nível superior de uma modalidade da invenção na qual um estimador de pdf é implementado no terminal receptor.

Para facilitar o entendimento, números de referência iguais forma usados, sempre que possível, para designar elementos idênticos que são comuns às figuras.

Descrição Detalhada da Invenção

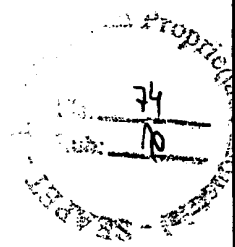
A invenção será discutida no contexto de uma rede de atraso variável como a Internet, em que a rede de atraso variável tende a comunicar erros temporais em pacotes de vídeo e/ou áudio transmitidos, de maneira tal que possa resultar em erros de sincronização dos lábios. A invenção opera em unidades de apresentação de vídeo e/ou áudio, tais como quadros de vídeo e áudio, unidades de apresentação estas que podem ser empacotadas para transporte adequado via uma rede, tal como uma rede de atraso variável.

Além disso, embora uma definição de comunicações padrão de "sincronização dos lábios" relacione sincronização (ou processo de sincronização) de fala ou canto com vídeo, de maneira tal que os movimentos dos lábios de vídeo pareçam coincidir naturalmente com o som; com propósitos da presente invenção, a definição não deve ser interpretada limitada desta maneira. Em vez disso, "sincronização dos lábios" refere-se à sincronização de qualquer ação representada em vídeo correspondente a uma trilha de áudio ou fluxo de bits, de maneira tal que o som supostamente gerado pela ação case devidamente com o vídeo que supostamente produz esse som. Em outras palavras, "sincronização dos lábios", com os propósi-



tos da presente invenção, diz respeito à sincronização entre som representados por um sinal de informação de áudio e vídeo correspondentes representados por um sinal de informação; independente do assunto objeto do áudio e vídeo correspondentes. Portanto a referência a um "erro de sincronização dos lábios" é de natureza geral e é interpretado como qualquer tipo de "erro de sincronização temporal audiovisual".

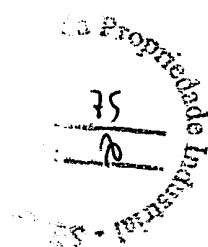
A Figura 1 descreve um diagrama de blocos de nível superior de um sistema de comunicações que inclui a presente invenção. Especificamente, o sistema de comunicações 100 compreende uma fonte audiovisual 110, tal como um dispositivo de armazenamento de massa, câmera, microfone, alimentação de rede ou outra fonte de informação audiovisual. A fonte audiovisual 110 fornece um fluxo de vídeo V a um codificador de vídeo 120V e um fluxo de áudio correspondente A a um codificador 120A, respectivamente. Os codificadores 120v e 120A, que ilustrativamente formam um MPEG ou outros codificadores de compactação, codificam o fluxo de vídeo V e fluxo de áudio A para produzir, respectivamente, fluxo de vídeo codificado VE e fluxo de áudio codificado AE. Os fluxos de vídeo VE e áudio AE codificados são processados por um processador de transporte 130 de acordo com o formato de transporte particular apropriado para a rede atrasada variável 140, ilustrativamente uma Ethernet, ATM ou outro codificador de fluxo de transporte, que codifica os fluxos de vídeo VE e áudio AE de acordo com o formato de transporte particular apropriado à rede de atraso variável 140.



O fluxo de transporte T é propagado por uma rede de atraso variável 140, tal como a Internet, intranet, ATN, Ethernet, LAN, WAN, rede telefônica comutada pública (PSTN), satélite ou outra rede; a um destino onde ele é recebido como fluxo de transporte T'. O fluxo de transporte T' compreende o fluxo de transporte original T que inclui qualquer atraso ou outros erros introduzidos por transporte pela rede de atraso variável 140.

O fluxo de transporte resultante T' é recebido por um processador de transporte 150, ilustrativamente uma Ethernet, ATM ou outro decodificador de fluxo de transporte, que extrai do fluxo de transporte recebido T' um fluxo de vídeo codificado VE' e um fluxo de áudio codificado correspondente AE'. Os fluxos de vídeo VE' e áudio AE' codificados compreendem fluxos de vídeo VE e áudio AE codificados iniciais que incluem qualquer erro, tais como erros temporais induzidos pelo processador de transporte 130, rede de atraso variável 140 e/ou processador de transporte 150. Os fluxos de vídeo VE' e áudio AE' codificados recebidos são decodificados por um decodificador 160 para produzir fluxos de vídeo V' e áudio A' resultantes. Os fluxos de vídeo V' e áudio A' resultantes são apresentados por um dispositivo de apresentação 170, tais como uma televisão ou outro dispositivo de exibição, 170V que tem associado com ele meios de apresentação de áudio, tais como alto-falantes 170A.

A Figura 2 descreve um diagrama de blocos de um controlador adequado para uso nos sistemas e aparelhos de acordo com os princípios da presente invenção. Especifica-



mente, o controlador 200 da figura 2 pode ser usado para implementar um ou mais dos elementos funcionais supradescritos com relação à figura 1, bem como os vários elementos funcionais descritos a seguir com relação às figuras 7 e 8.

5 O controlador exemplar 200 da figura 2 compreende um processador 230 bem como uma memória 240 para armazenar vários programas 245. O processador 230 coopera com componentes de circuito de suporte convencionais 220, tais como fontes de potência, circuitos de relógio, memória de provi-

10 são e similares, bem como circuitos que ajudam na execução da rotina de suporte lógico armazenada na memória 240. Como tal, considera-se que algumas das etapas de processo aqui discutidas como processos de suporte lógico possam ser implementadas em hardware, por exemplo, como componentes de

15 circuito que cooperam com o processador 230 para realizar várias etapas. O controlador 200 também contém componentes de circuito de entrada/saída (I/O) 210 que formam uma interface entre os vários elementos funcionais que se comunicam com um elemento funcional que inclui o controlador 200.

20 Embora o controlador 200 da figura 2 esteja representado como um computador de uso geral que é programado para realizar várias modificações temporais de fluxos de áudio e/ou vídeo de acordo com a presente invenção, a invenção pode ser implementada em hardware, por exemplo como um circui-

25 to integrado de aplicação específica (ASCI). Como tal, as etapas de processo aqui descritas devem ser interpretadas de forma geral como realizadas equivalentemente por suporte lógico, hardware ou uma combinação destes.

O Erro de Sincronização dos lábios (LSE) pode ser definido de acordo com a equação 1, como se segue:

$$e = (t_d^a - t_d^v) - (t_e^a - t_e^v) \quad (\text{eq. 1})$$

Na equação 1, t_d^a e t_e^v são o tempo em que os quadros de áudio e vídeo relacionados chegam no dispositivo de apresentação 170 no lado do receptor, respectivamente; e t_e^a e t_e^v são o tempo em que os quadros de áudio e vídeo chegam aos codificadores de áudio e vídeo, respectivamente.

A Figura 3 descreve uma representação gráfica de uma função de densidade de probabilidades $p(e)$ de um erro de sincronização dos lábios e útil no entendimento da presente invenção. Por causa do atraso randômico introduzido, por exemplo, pela rede de atraso variável 140, o atraso entre o pacote de áudio e seu pacote de vídeo correspondente no lado do receptor é uma variável aleatória. Esta variável aleatória é confinada pela sua função densidade de probabilidades (pdf), $p(e)$, como a linha cheia 310 da figura 3. Especificamente, a representação gráfica da figura 3 descreve um eixo geométrico horizontal que define um relacionamento temporal entre dados de vídeo e dados de áudio correspondentes. Um tempo zero é selecionado para representar um tempo no qual os dados de vídeo representam conteúdo que tem associado com ele dados de áudio sincronizados. Embora esta distribuição esteja representada como uma distribuição Gaussiana, outras curvas pdf simétricas ou assimétricas podem ser utilizadas, dependendo da fonte de erro particular modelada, bem como do número de fontes de erro modeladas (isto é, uma curva pdf simétrica ou assimétrica compostas por múltiplas fontes de



vídeo e áudio pode ser usada).

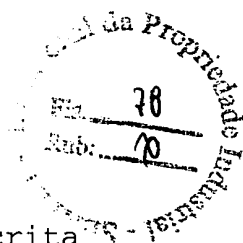
À medida em que o tempo aumenta de zero na direção positiva, diz-se que os dados de áudio ficam cada vez mais para trás dos dados de vídeo (isto é, pacotes de áudio ficam cada vez mais atrasados em relação aos pacotes de vídeo correspondentes). À medida em que o tempo aumenta na direção negativa em relação ao zero, diz-se que os dados de áudio ficam cada vez mais para frente dos dados de vídeo (isto é, os pacotes de vídeo são cada vez mais atrasados em relação aos pacotes de áudio correspondentes).

A Figura 4 descreve uma representação gráfica de uma janela de tolerância de erro de sincronização dos lábios (LSET) útil no entendimento da presente invenção. Especificamente, a janela LSET é definida pela função da equação (2), como se segue, onde a e b são os limites inferior e superior da tolerância da janela LSET.

$$W(e) = \begin{cases} 1(a \leq e \leq b) \\ 0(\text{de outra forma}) \end{cases} \quad (\text{eq. 2})$$

Os inventores notaram que as tolerâncias de erro assimétricas para pacotes de áudio e vídeo e inúmeros problemas que surgem de casos em que um pacote de áudio é recebido antes do pacote de vídeo correspondente. Uma faixa típica dos valores varia, por exemplo, $[a, b] = [20 \text{ ms}, 40 \text{ ms}]$.

A Figura 5 descreve uma representação gráfica de um deslocamento da pdf dentro de uma janela de tolerância. Especificamente, a representação gráfica da figura 5 descreve um eixo geométrico horizontal que define um relacionamento temporal entre dados de vídeo e dados de áudio correspon-



dentes. Um tempo zero é selecionado da maneira supradescrita em relação às figuras 3 e 4. A janela de tolerância de atraso 410 descreve a tolerância de atraso ou erros temporais associados com o sincronismo dos lábios que não serão questionáveis por um espectador. Nota-se que a janela de tolerância de atraso 410 da figura 5 estende-se de -20 milissegundos (isto é, os pacotes de áudio adiantando os pacotes de vídeo em até 40 milissegundos) até +40 milissegundos (isto é, os pacotes de áudio atrasando os pacote de vídeo em até 40 milissegundos). Nota-se que erros de sincronização dos lábios em que a informação de áudio avança a informação de vídeo tende ser mais questionável (por exemplo, mas perceptível e/ou incômodo para um observador) do que no caso em que a informação de áudio atraso a informação de vídeo, assim a assimetria na janela de tolerância de atraso 410 da figura 5.

Referindo-se à figura 5, a parte da "cauda" esquerda de uma curva pdf 510 cai em uma região 540 além de uma faixa da janela de tolerância de atraso inferior. Nota-se que a parte da "cauda" direita da curva pdf 510 é substancialmente zero bem antes da extremidade superior da faixa de tolerância da janela de atraso. A faixa de tolerância da janela de erro é definida como a faixa na qual erros temporais tais como erros de sincronização dos lábios são considerados menos ofensivos. Assim, atrasos que excedem tanto positiva como negativamente a faixa de tolerância de atraso (isto é, atrasos fora de uma janela de tolerância de erros) compreende esses atrasos que são considerados questionáveis

ou altamente questionáveis pela média dos observadores.

Uma curva pdf deslocada 520 descreve a curva de distribuição de probabilidades inicial 510 deslocada com o tempo, de maneira tal que uma maior área abaixo da curva pdf fique dentro da janela de tolerância de erros 410. Assim, a primeira pdf, ou inicial, foi deslocada no tempo de maneira tal que uma maior área (preferivelmente uma área máxima) sob a segunda pdf, ou final, seja incluída na janela de tolerância de erro 410. Este deslocamento em pdf é causado pela adaptação de parâmetro(s) de sincronismo associado(s) com 10 formação de vídeo e/ou áudio, tais como carimbos temporais de apresentação de unidades de acesso de vídeo e/ou áudio. Assim, se informação de áudio e/ou vídeo temporais for adaptada para realizar um deslocamento como esse na pdf correspondente, então a probabilidade de erros de sincronização 15 dos lábios questionáveis é minimizada, ou pelo menos reduzida, em uma quantidade mensurável com a redução de pdf sob o erro da curva causada pelo deslocamento. Portanto, a solução ideal para maximização da área sob a curva LSE dentro do LSET é maximizar a função objeto dada como a equação 3, da 20 maneira seguinte:

$$\begin{aligned}
 J &= \int_{-\infty}^{+\infty} p(e-t_0)W(e)de \\
 &= \int_a^b p(e-t_0)de \\
 &= P(b-t_0) - P(a-t_0)
 \end{aligned}
 \tag{eq. 3}$$

Na equação 3, $p(e)$ é a pdf da LSE, $P(e)$ é a função de distribuição cumulativa e $W(e)$ é a função da janela LSET 25 definida em [2], respectivamente. O processo de otimização é maximizar a área encerrada pela curva pdf ligada por $[a, b]$.

Isto é equivalente ao processo de minimização da área "da cauda" fora da janela. Este problema de otimização pode ser solucionado derivando-se J em relação a t_0 e resolvendo-se a equação 4 para t_0 como se segue:

5
$$\frac{dJ}{dt_0} = 0 \quad (\text{eq. 4})$$

Pode-se provar que a solução ideal de t_0 para uma pdf LSE Gaussiana mostrada na figura 2 é a média dos limites inferior e superior da janela LSET:

$$t_0 = \frac{a+b}{2} \quad (\text{eq. 5})$$

10 Para a outra pdf LSE, o t_0 ideal pode ter um valor positivo ou negativo, dependendo das localizações geográficas relativas entre a pdf e a janela de tolerância de erro. Um t_0 positivo significa atrasos nos quadros de áudio e atrasos de t_0 negativos nos quadros de vídeo para deslocar o
15 LSE e maximizar a equação 4.

A Figura 6 descreve um método para processar pacotes de áudio e/ou vídeo de acordo com a invenção. Especificamente, a figura 6 descreve um método para adaptar quadro de vídeo e/ou áudio correspondentes ou acessar pacotes de
20 unidade de maneira tal que erros de sincronização dos lábios e, mais particularmente, erros de sincronização dos lábios do tipo antecipação áudio sejam minimizados. No contexto do método da figura 6, erros de sincronização dos lábios são induzidos, pela caixa 605, por uma fonte de erros que com-
25 preende uma ou mais redes de atraso variável, um codificador, um processador de transporte ou outra fonte de erro.

Na etapa 610, os erros temporais que provavelmente são produzidos por uma fonte de erros são representados como uma função de densidade de probabilidade (pdf). Por exemplo, conforme supradescrito com relação à figura 5, está mostrada uma pdf associada com erros temporais prováveis de ser induzidos por uma rede de atraso variável. Esta pdf compreende, ilustrativamente, uma distribuição de números aleatórios que tem uma forma Gaussiana (pode ou não ser centralizada em zero), onde zero descreve nenhum erro de sincronização dos lábios (isto é, alinhamento temporal de dados de vídeo e áudio).

Na etapa 620, uma janela de tolerância de erro associada com a pdf é definida. Conforme notado na caixa 615, a janela de tolerância de erro pode ser definida em relação ao erro de sincronização dos lábios ou outros erros. Conforme notado na figura 5, uma janela de tolerância de atraso associada com erros de sincronização dos lábios é definida, ilustrativamente, como os atrasos entre -20 milissegundos e +40 milissegundos. Ou seja, um valor de tolerância de atraso de áudio assimétrico (em relação ao ponto de tempo zero) é provido com unidades de acesso de áudio que avança unidades de acesso de vídeo em até 20 milissegundos, que atrasa pacotes de vídeo correspondentes em até 40 milissegundos é considerado tolerável. Outras janelas de tolerância de atraso podem ser definidas, dependendo dos fatores associados com um sistema de comunicações que utiliza a presente invenção.

Na etapa 630, o método adapta parâmetros de sincronismo tais como carimbos temporais associados com pelo menos um dos quadros de vídeo e áudio que formam um fluxo de

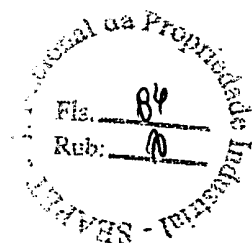


conteúdo. Opcionalmente, um ou ambos fluxos de áudio e vídeo não compactados são atrasados antes da codificação. Esta adaptação é feita de uma maneira a tender provocar um deslocamento na pdf associada com a fonte de erros de uma posição inicial (por exemplo, centrada em torno do zero) em direção a uma posição máxima utilizando a janela de tolerância de atraso. Nota-se, na caixa 625, que tal adaptação pode ocorrer durante um processo de codificação, um processo de transporte ou outro processo. Referindo-se novamente à figura 5, um deslocamento de pdf apropriado está mostrado como um que aumenta a quantidade de área sob a curva de distribuição de probabilidades que está dentro dos limites estabelecidos pela janela de tolerância de atraso.

A Figura 7 descreve um diagrama de blocos de nível superior de um sistema de comunicação de acordo com uma modalidade alternativa da invenção. Especificamente, o sistema de comunicação 700 da figura 7 é substancialmente o mesmo do sistema de comunicação 100 da figura 1. A principal diferença é que o elemento de atraso 710A é usado para atrasar o fluxo de áudio inicial A antes da codificação deste fluxo de áudio pelo codificador de áudio 120A. O elemento de atraso 710A comunica um atraso de t_0 ao fluxo de áudio de forma a deslocar uma pdf correspondente de acordo com o modelo de tolerância de erro de sincronização dos lábios (LSET) supradiscutido. Nota-se que o sistema de comunicação 700 da figura 7 pode ser modificado de forma a incluir um elemento de atraso de vídeo correspondente 710V (não mostrado) para atrasar o sinal de fonte de vídeo V antes da codificação pelo

Nesta modalidade da invenção, no caso de a janela de tolerância de erro 410 mostrada na figura 5 ser utilizada, cada qual, ilustrativamente, unidade de quadros de áudio ou de acesso é atrasada aproximadamente t_0 milissegundos em relação a cada quadro de vídeo antes da codificação. Deslocando-se cada quadro de áudio para trás no tempo t_0 milissegundos, a pdf associada com os erros induzidos pela rede de atraso variável é deslocada da maneira supradescrita com relação à figura 5. Ou seja, a pdf é deslocada para frente ou para trás, dependendo do sinal de t_0 , no tempo de uma tendência de ter um erro de sincronização dos lábios do pacote de áudio com antecedência para uma tendência de nenhum erro de sincronização ou erro de sincronização dos lábios do pacote de áudio com atraso (que é menos questionável do que um erro de sincronização dos lábios do pacote de áudio com antecedência). Assim, aumenta-se a probabilidade de que nenhum atraso do pacote de áudio permaneça dentro dos limites de tolerância de erro estabelecida pela janela de tolerância de erro 410.

Em uma modalidade da invenção, no caso de uma pdf Gaussiana simétrica, tal como mostrada na figura 2, ser considerada, os carimbos temporais dos quadros de áudio ou vídeo são modificados para minimizar os desencontros de sincronização. Para um codificador de áudio com taxa de bit constante, o carimbo temporal de vídeo é opcionalmente modificado de uma maneira a tender a aumentar a probabilidade do



desacordo de sincronismo permanecer dentro do LSET no lado do decodificador. Nesta modalidade, ilustrativamente, os carimbos temporais de vídeo são arredondados para os décimos mais baixos de milissegundos, tal como indicado pela equação 6, a seguir (onde t_t^v e t_e^v são carimbos temporais original e arredondado para um quadro de vídeo em milissegundo):

$$\hat{t}_e^v = t_e^v - (t_e^v \bmod 10) \quad (\text{eq. 6})$$

A técnica citada introduz um atraso uniformemente distribuído nos pacotes de áudio na faixa de 0 a 9 milissegundos. Outras faixas podem ser selecionadas (por exemplo, mod 15, mod 20, etc.) e os pacotes de áudio podem também ser processados desta maneira.

Nas modalidades previamente descritas, as pds de LSE são conhecidas e presumidas como ligeiramente estáveis. Em decorrência disto, um deslocamento de tempo predeterminado é realizado em todas as unidades de acesso de áudio (ou vídeo). Em uma modalidade mais avançada, em que pdf LSE pode não ser conhecida, ou que não seja estável, a pdf LSE é monitorada e estimada, e o deslocamento de tempo não é predeterminado.

A Figura 8 descreve o LSE na modalidade em que o estimador de pdf é implementado no lado do receptor. Especificamente, a aparelho no lado do receptor, tal como representado anteriormente nas figuras 1 e 7 é modificado de forma a incluir um estimador de pdf LSE 810 e um elemento de atraso de áudio 820A. Embora não mostrado, um elemento de atraso de vídeo 820V pode também ser utilizado. O estimador pdf de LSE 810 recebe os sinais de áudio A' e de vídeo V'



decodificados e, em resposta à informação do modelo de LSET, produz um sinal indicativo de atraso t_0 . Na modalidade da figura 8, o sinal indicativo do atraso t_0 é processado pelo elemento de atraso de áudio 820A para comunicar uma quantidade correspondente de atraso no fluxo de áudio decodificado A', produzindo assim um fluxo de áudio atrasado A". O estimador 810 coleta constantemente carimbos temporais de apresentação das unidades de acesso de áudio e vídeo. Cada LSE é calculado com uso da equação 1. Todos LSE são usados para formar a pdf do LSE. Com uso do modelo de LSET, o deslocamento de tempo ideal t_0 pode ser derivada, solucionando-se a equação 4 para o deslocamento de tempo t_0 . O atraso tanto no quadro de áudio ($t_0 > 0$) como no quadro de vídeo ($t_0 < 0$) é adicionado à deslocamento na pdf de LSE.

Em uma modalidade, o deslocamento de tempo ótico determinado propaga-se do receptor até o codificador, de maneira tal que pelo menos um dos fluxos de áudio e vídeo a ser codificado e transmitido seja atrasado antes da codificação, antes do processamento do transporte e/ou antes do transporte ao receptor.

Embora várias modalidade que incorporam os preceitos da presente invenção tenham sido mostrados e descritos aqui com detalhes, técnicos habilitados podem facilmente perceber muitas outras modalidades variadas que ainda incorporam esses preceitos.

REIVINDICAÇÕES

1. Método, **CARACTERIZADO** pelo fato de que compreende:

receber (610) unidades de acesso de vídeo e unidades de acesso de áudio correspondentes, as ditas unidades de acesso de vídeo e áudio correspondentes representando informação audiovisual que tende a apresentar um erro de sincronização temporal audiovisual descrita por uma primeira função de distribuição de probabilidade (pdf) (510); e

atrasar temporalmente (630) uma das unidade de acesso de áudio e vídeo recebidas por unidades de acesso de quantidade de atraso, não atrasadas e atrasadas correspondentes representando informação audiovisual que tende a apresentar um erro de sincronização temporal audiovisual descrito por uma segunda pdf (520), a dita segunda pdf utilizando uma parte maior de uma janela de tolerância de sincronização (410) do que a dita primeira pdf.

2. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que:

a dita primeira pdf tem associada com ela um valor de desacordo temporal negativo e um valor de desacordo temporal positivo, os ditos valores de desacordo temporal tendo diferentes valores absolutos; e

a dita quantidade de atraso sendo selecionada para reduzir a diferença em valor absoluto dos ditos valores de desacordo temporal negativo e positivo (630).

3. Método, de acordo com a reivindicação 2, **CARACTERIZADO** pelo fato de que o dito valor de desacordo

temporal negativo é de 20 milissegundos, o dito valor de desacordo temporal positivo é de 40 milissegundos, e as ditas unidades de acesso de áudio são atrasadas em 10 milissegundos.

5 4. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que a dita quantidade de atraso é comunicada antes de uma codificação de fluxos de áudio e vídeo que produz as ditas unidades de acesso de áudio e vídeo (710).

10 5. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que a dita quantidade de atraso é comunicada durante uma codificação de fluxos de áudio e vídeo que produz as ditas unidades de acesso de áudio e vídeo.

 6. Método, de acordo com a reivindicação 1, **15 CARACTERIZADO** adicionalmente pelo fato de que compreende: adaptar (810, 820) a dita quantidade de atraso em resposta às mudanças na dita primeira pdf.

 7. Método, de acordo com a reivindicação 6, **20 CARACTERIZADO** pelo fato de que as ditas mudanças são determinadas examinando-se carimbos temporais de apresentação das unidades de acesso de áudio e vídeo em um receptor.

 8. Método, de acordo com a reivindicação 1, **25 CARACTERIZADO** adicionalmente pelo fato de que compreende: adaptar a dita quantidade de atraso em resposta às mudanças em uma fonte de indução de erro de sincronização temporal audiovisual.

 9. Método, de acordo com a reivindicação 8, **CARACTERIZADO** pelo fato de que a dita fonte de indução de

erro de sincronização temporal audiovisual compreende uma rede de atraso variável (140).

10. Método para produzir fluxos de vídeo codificado e de áudio correspondentes, compreendendo:

5 codificar (120) informação de vídeo e áudio temporalmente correspondentes para produzir fluxos de vídeo e áudio codificados compreendendo as respectivas unidades de acesso de vídeo e áudio; e

CARACTERIZADO por o método ainda compreender

10 atrasar temporalmente (710) um dos ditos fluxos de vídeo e áudio codificados por uma quantidade de atraso temporal correspondente a um modelo de tolerância de erro,

em que o dito modelo de tolerância de erro define uma janela de tolerância de sincronização (410) e o dito
15 atraso temporal faz com que uma função de distribuição de probabilidades (pdf) (310) que descreve um erro de sincronização entre o dito sinal de áudio e o sinal de vídeo correspondente seja deslocado em direção a uma correspondência mais favorável com a dita janela de tolerância de sincroni-
20 zação (410).

11. Método, de acordo com a reivindicação 10, **CARACTERIZADO** pelo fato de que a dita etapa de atrasar temporalmente compreende atrasar um dos ditos fluxos de vídeo e áudio antes da dita codificação.

25 12. Método, de acordo com a reivindicação 10, **CARACTERIZADO** pelo fato de que cada um dos ditos fluxos de vídeo e áudio codificados compreende uma pluralidade de respectivas unidades de acesso de vídeo e áudio; e

a dita etapa de atrasar temporalmente compreende adaptar carimbos temporais associados com pelo menos uma das ditas unidades de acesso de vídeo e áudio pela dita quantidade de atraso.

5 13. Método, de acordo com a reivindicação 10, **CARACTERIZADO** pelo fato de que carimbos temporais associados com um dos ditos fluxos de vídeo e áudio codificados são arredondados por uma quantidade predeterminada.

10 14. Método, de acordo com a reivindicação 13, **CARACTERIZADO** pelo fato de que os ditos carimbos temporais de fluxo de vídeo são estimados reduzindo-se cada carimbo temporal de fluxo de vídeo em um respectivo carimbo temporal de fluxo de vídeo respectivo como processado por uma operação de módulo.

15 15. Método, de acordo com a reivindicação 10, **CARACTERIZADO** pelo fato de que a quantidade de atraso temporal é adaptada em resposta às mudanças em uma fonte de produção de erro de sincronização temporal audiovisual.

20 16. Aparelho que compreende:
um codificador (120) para codificar sinais de áudio e vídeo para produzir fluxos de áudio e vídeo codificados; e

CARACTERIZADO por o aparelho ainda compreender
um elemento de atraso (710) para comunicar um
25 atraso temporal em pelo menos um do sinal de áudio e do sinal de vídeo correspondente em resposta a um modelo de tolerância de erro,

em que o dito modelo de tolerância de erro define

uma janela de tolerância de sincronização (410) e o dito atraso temporal faz com que uma função de distribuição de probabilidades (pdf) (310) que descreve um erro de sincronização entre o dito sinal de áudio e o sinal de vídeo correspondente seja deslocado em direção a uma correspondência mais favorável com a dita janela de tolerância de sincronização (410).

17. Aparelho, de acordo com a reivindicação 16, **CARACTERIZADO** pelo fato de que a dita janela de tolerância de sincronização (410) tem associada com ela um valor de desacordo temporal negativo e um valor de desacordo temporal positivo, os ditos valores de desacordo temporais possuindo diferentes valores absolutos, a dita pdf possuindo respectivos valores de desacordo temporal negativo e positivo que são deslocados em direção ao alinhamento com os ditos valores de desacordo temporal da janela de tolerância de sincronização.

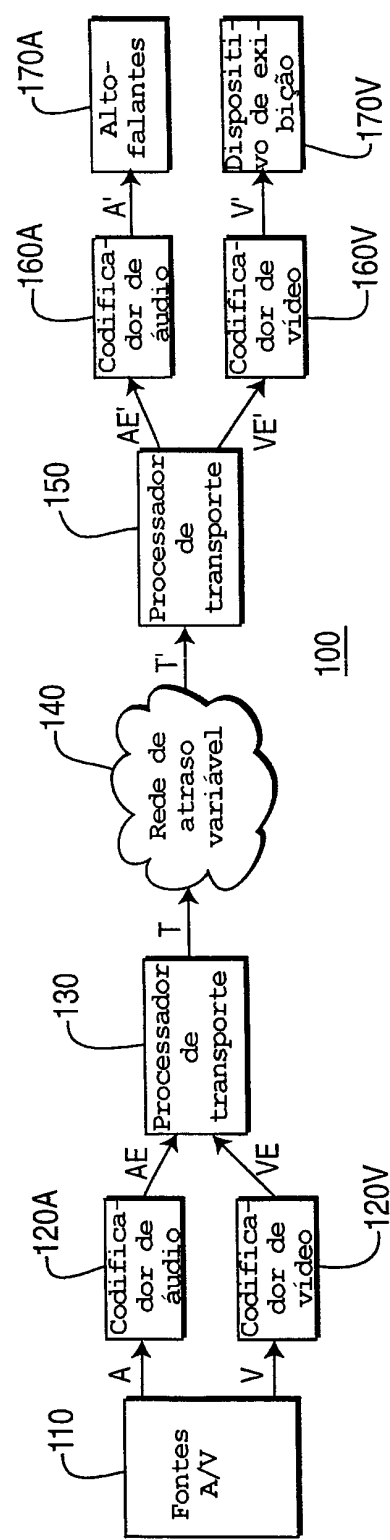


FIG. 1

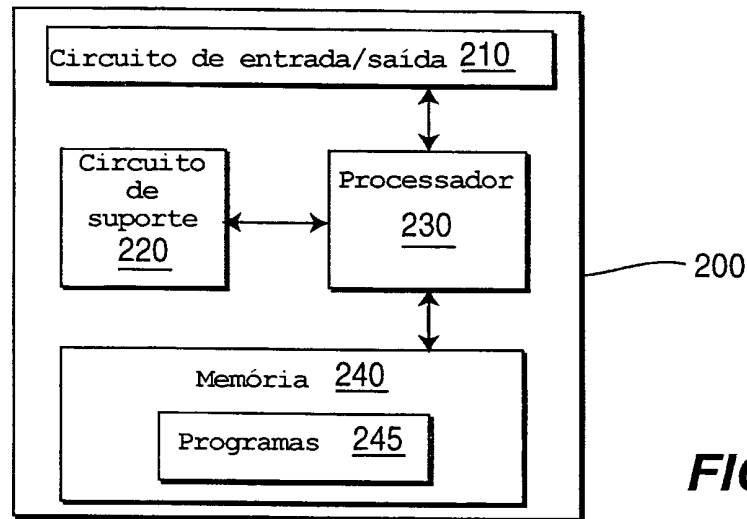
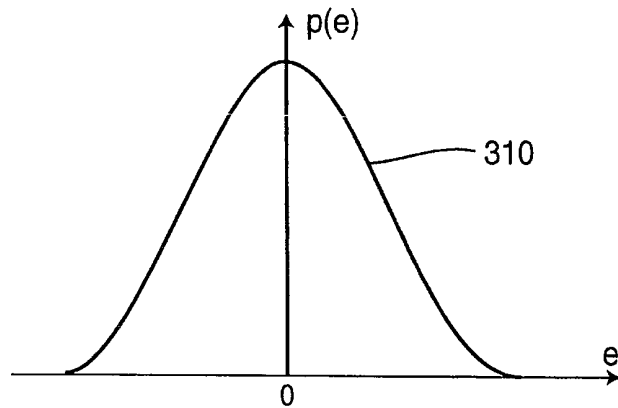
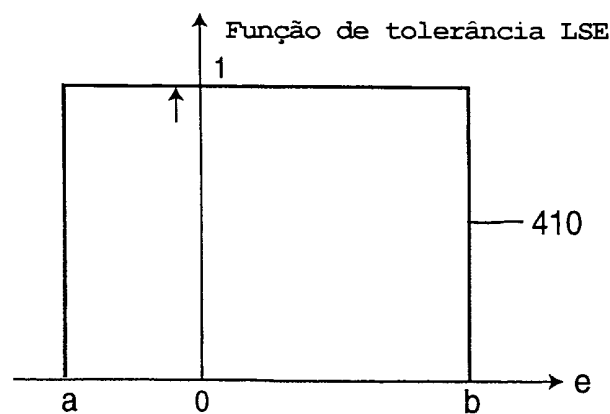


FIG. 2



Probabilidade da função de densidade $p(e)$ do erro de sincronização dos lábios e .

FIG. 3



Janela de tolerância de erro de sincronização dos lábios (LSET).

FIG. 4

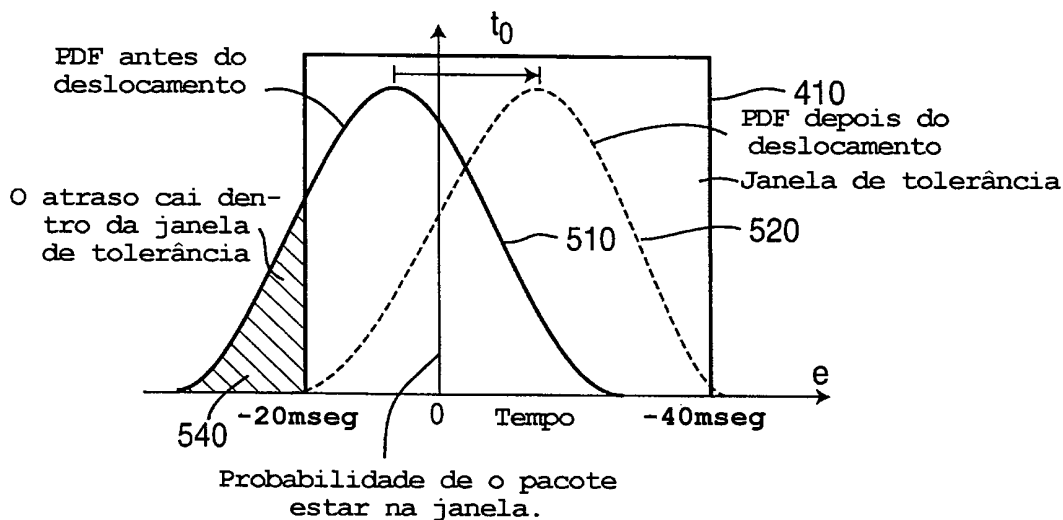


FIG. 5

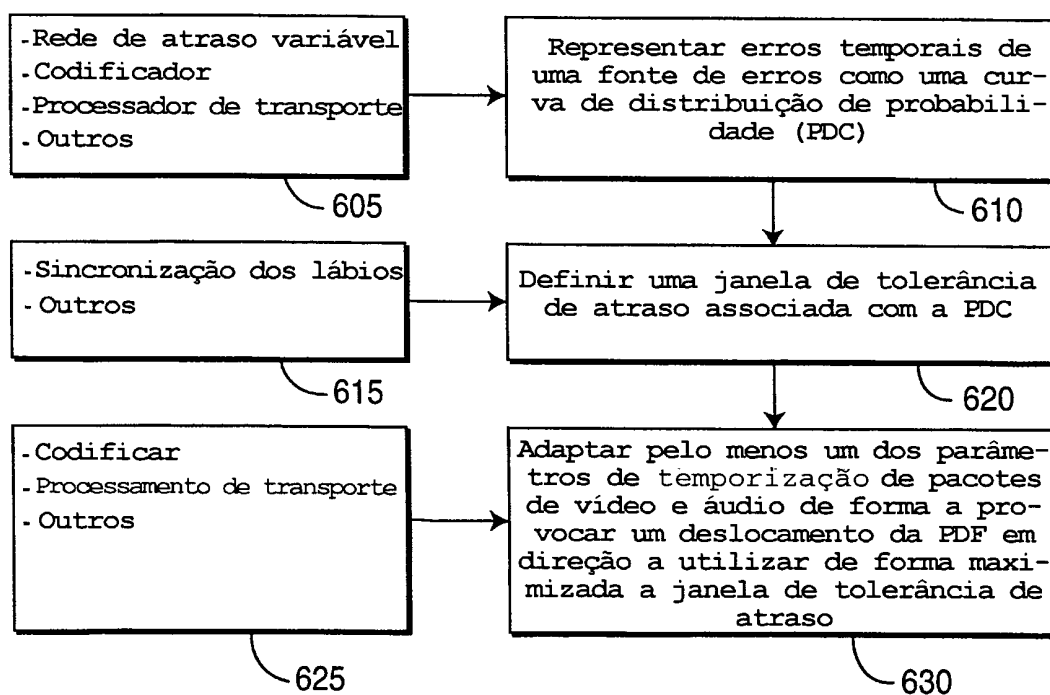


FIG. 6

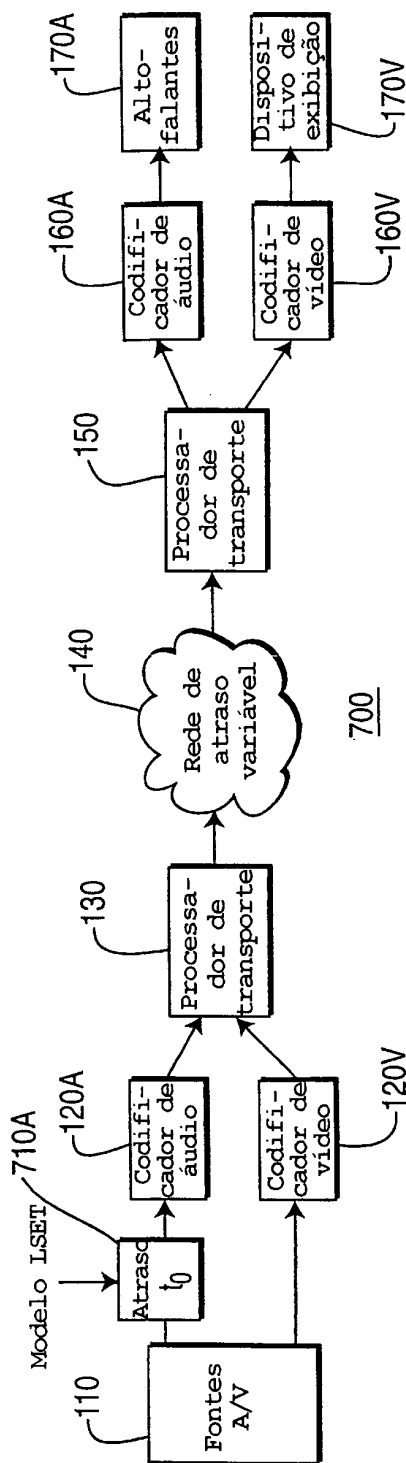


FIG. 7

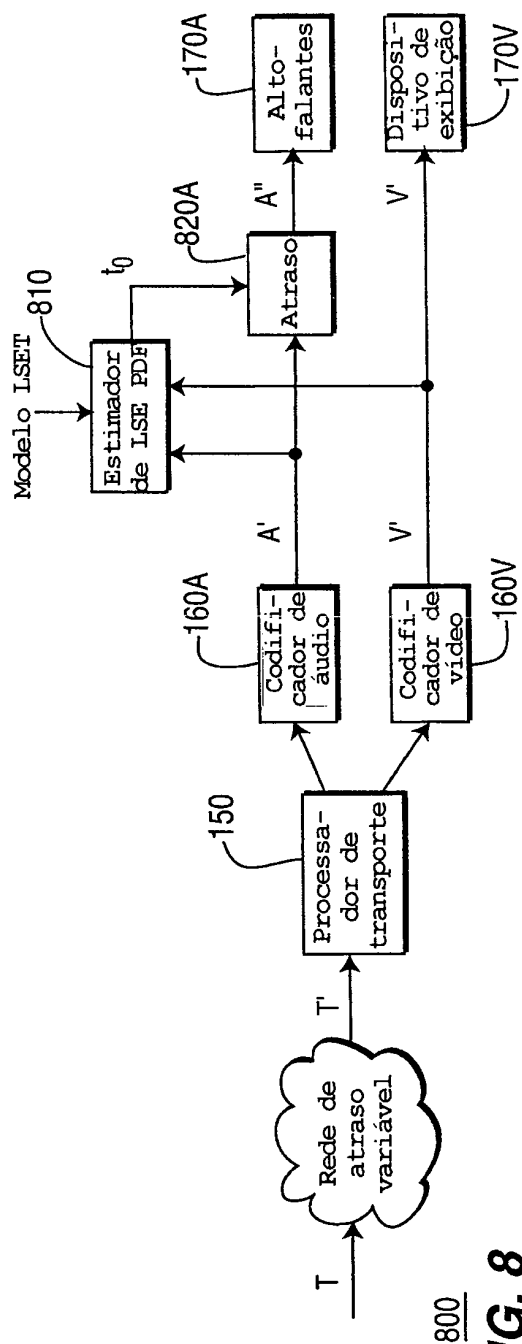


FIG. 8